Pattern Recognition Letters
journal homepage: www.elsevier.com

EI3D: Expression-Invariant 3D Face Recognition based on Feature and Shape Matching

Yulan Guo[a,**], Yinjie Lei[b], Li Liu[c], Yan Wang[d], Mohammed Bennamoun[e], Ferdous Sohel[f]

[a] *College of Electronic Science and Engineering, National University of Defense Technology, Changsha, China*
[b] *College of Electronics and Information Engineering, Sichuan University, Chengdu, Sichuan, China*
[c] *College of Information System and Management, National University of Defense Technology, Changsha, China*
[d] *College of Computer Science, Sichuan University, Chengdu, Sichuan, China*
[e] *School of Computer Science and Software Engineering, The University of Western Australia, WA, Australia*
[f] *School of Engineering and Information Technology, Murdoch University, WA, Australia*

ABSTRACT

This paper presents a local feature based shape matching algorithm for expression-invariant 3D face recognition. Each 3D face is first automatically detected from a raw 3D data and normalized to achieve pose invariance. The 3D face is then represented by a set of keypoints and their associated local feature descriptors to achieve robustness to expression variations. During face recognition, a probe face is compared against each gallery face using both local feature matching and 3D point cloud registration. The number of feature matches, the average distance of matched features, and the number of closest point pairs after registration are used to measure the similarity between two 3D faces. These similarity metrics are then fused to obtain the final results. The proposed algorithm has been tested on the FRGC v2 benchmark and a high recognition performance has been achieved. It obtained the state-of-the-art results by achieving an overall Rank-1 recognition rate of 97.0 percent and an average verification rate of 99.01 percent at 0.001 false acceptance rate for all faces with neutral and non-neutral expressions.

1. Introduction

The human face is considered to be one of the most important biometrics due to its high accessibility, social acceptability, and non-intrusiveness nature [31, 25]. Facial biometrics has a number of applications including surveillance, security, entertainment, commerce, and forensics [3, 39]. It is particularly suitable for applications where other biometrics (including iris images, retinal scans and fingerprints) are not available or desirable [30].

Face recognition can be performed using 2D facial images, 3D facial scans or their combination [3, 27]. 2D face recognition has been extensively investigated during the past few decades [48, 35, 28]. However, 2D face recognition is still challenged by a number of factors including illumination variations, scale differences, pose changes, facial expressions and makeup. Moreover, affine transformations are introduced to 2D images during acquisition, which make 2D face recognition even more difficult [6, 5]. With the rapid development of 3D scanners, 3D data acquisition is becoming increasingly cheaper and non-intrusive [2]. Besides, 3D facial scans are more robust to lighting conditions, pose variations and facial makeup [6]. The 3D geometry represented by a facial scan also provides a new clue for accurate face recognition. 3D face recognition is therefore believed to have the potential to overcome many limitations encountered by 2D face recognition [9], and has been considered as an alternative or complementary solution to conventional 2D face recognition approaches [6, 25].

Several approaches have been proposed to address various aspects of a 3D face recognition system. According to the facial representation types, the existing work can be classified into landmark-based, curve-based, patch-based, and holistic algorithms [3]. The landmark-based algorithms represent each face with a set of local features by calculating the relations (e.g., distances and angles) between a set of facial landmarks (fiducial points), examples include the anthropometric facial distance feature [22, 17, 34, 46]. The curve-based algorithms represent each face with a set of curves, including iso-depth curves, iso-radius curves, iso-geodesic curves and pro-

**Corresponding author
e-mail: yulan.guo@nudt.edu.cn (Yulan Guo)

files [40, 38, 4, 42, 26, 8]. The curve-based algorithms encode more geometric information of the 3D facial surface and are believed to be more discriminative than the landmark-based algorithms [40]. The patch-based algorithms first extract several patches from a 3D facial surface and then encode the geometric information of each patch with a feature descriptor. Examples of patch features include point signatures [12, 13], 3D geometric features [24], Mesh Scale-Invariant Feature Transform (Mesh-SIFT) [41], and Mesh Histogram of Oriented Gradients (Mesh-HOG) [47, 5]. The holistic algorithms use the information of the entire face or large regions of the 3D face to perform face recognition. Examples include Iterated Closest Point (ICP) based surface matching algorithms [30, 33], extended Gaussian images [45, 43], canonical forms [11], spherical harmonic features [29], and the tensor representation [1]. A major limitation of the holistic algorithms is that they require accurate normalization of the 3D faces, and they are commonly more sensitive to facial expressions and occlusions [31]. For a comprehensive review on existing 3D face recognition algorithms, the reader should refer to [40, 9, 10, 3].

For a face recognition system, high recognition accuracy and a strong robustness are the two key considerations for many practical applications [2]. Facial expression variation is one of the major problems for face recognition, since the drastic and complex geometric deformation of a human face caused by facial expressions can dramatically deteriorate the recognition performance. Although the overall 3D shape of the face will be deformed by facial expression dynamics, the shape of several local facial surfaces (e.g., nose) can be well preserved. Compared to the holistic algorithms, the local feature based algorithms are more robust to various nuisances [31, 19].

Motivated by these considerations, we propose a fully automatic and expression-invariant 3D face recognition algorithm (called EI3D). The EI3D algorithm first detects the nosetip and the 3D face is then cropped and normalized. A set of class-specific keypoints are subsequently detected from each face. The distribution of keypoints varies among individuals and is highly related to the specific shape of the 3D face. Next, the local surface around each keypoint is represented with a Rotational Projection Statistics (RoPS) feature. Face recognition is finally achieved using both RoPS feature matching and face registration. The EI3D algorithm was tested on the Face Recognition Great Challenge (FRGC) dataset. It achieved high verification rates at 0.1% False Acceptance Rate (FAR) of 99.9% and 97.12% for probe faces with neutral and non-neutral expressions, respectively. It also achieved high Rank-1 identification rates of 99.4% and 94.0% for probe faces with neutral and non-neutral expressions, respectively.

The paper is organized as follows. Section 2 describes the preprocessing and keypoint detection approach, followed by 3D feature description and matching scheme in Section 3. Section 4 introduces the 3D face recognition approach. Section 5 gives the experimental results and analyses. Finally, Section 6 concludes this paper.

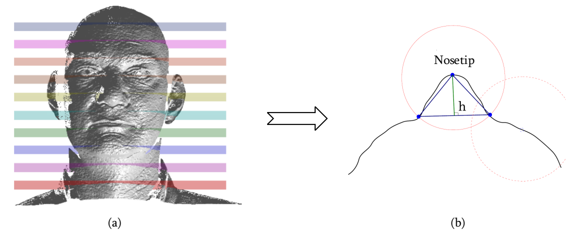


Fig. 2: An illustration of 3D facial nosetip detection. (a) Horizontal planes for 3D facial scan slicing. (b) Horizontal facial profile. (Figure best seen in color.)

2. Preprocessing and Keypoint Detection

2.1. Facial Data Preprocessing

Due to the characteristics of 3D sensors, a raw 3D facial scan acquired by a 3D sensor may suffer from many nuisances including missing data, spikes, and small pose variations [25]. Therefore, the raw facial scan have to be preprocessed before any further operations. Several preprocessing approaches can be found in the literature [14, 30, 24, 25, 36, 3]. In this paper, we use the work proposed in [30] to perform facial data preprocessing. The method consists of three parts, i.e., nosetip detection and face cropping, spike removal and hole filling, and pose normalization and resampling. An illustration of 3D facial data preprocessing is shown in Fig. 1.

2.1.1. Nosetip Detection and Face Segmentation

Given a raw facial scan acquired from the shoulder level up (as shown in Fig. 1(a)), we first detect the nosetip to remove undesired points outside the 3D facial region (as shown in Fig. 1(b)) [30]. First, a set of horizontal planes are used to slice the 3D facial scan, resulting in a set of horizontal profiles of the 3D face, as shown in Fig. 2(a). For each horizontal profile, the points on that profile are then uniformly interpolated to fill in holes. Then, a set of probe points are located on each profile and a circle is placed at each point, resulting in two intersection points with the horizontal profile, as shown in Fig. 2(b). A triangle is formed by the probe point and the two intersection points. The probe point with the largest altitude h of its associated triangle along the profile is considered to be a nosetip candidate. This process is repeated for all horizontal planes to obtain a set of nosetip candidates. These candidates are then refined using the Random Sample Consensus (RANSAC) algorithm [15]. The remaining candidates can then be considered lying on the noise ridge and the one with the largest altitude is considered to be the noisetip. Once the nosetip is detected, a 3D face is then cropped from the facial scan by eliminating the points which are located more than 80mm from the nosetip (as shown in Fig. 2(c)).

2.1.2. Spike Removal and Hole Filling

Once the 3D face is cropped from the facial scan, spikes are then removed by eliminating outlier points. Once the spikes are removed, the 3D face is uniformly resampled on the xy plane with a square grid resolution of 1mm. However, the spike removal process will result in undesired holes on the 3D face.

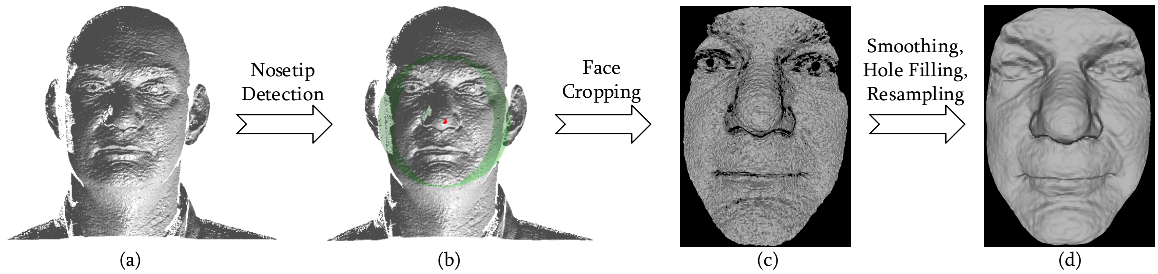


Fig. 1: An illustration of 3D facial data preprocessing. (a) Raw facial scan. (b) Facial scan with detected nosetip. (c) Cropped 3D face. (d) 3D face after spike removal, hole filling and smoothing.

Besides, the holes can also result from other factors including a light absorption in the dark areas, specular reflection of the underlying surface (e.g., the sclera, the pupil and the eyelashes), open mouths, and self-occlusion [3]. For 3D faces, these holes can be filled using cubic interpolation. Finally, noise is further smoothed using a median filter (as shown in Fig. 2(d)).

2.1.3. Pose Normalization and Resampling

Given the point cloud $\boldsymbol{P} = \left\{\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_{N_p}\right\} \in \mathbb{R}^3$ of a 3D face, where N_p is the number of points of the 3D face, the Hotelling transform is used to perform pose normalization [30]. First, the mean $\bar{\boldsymbol{p}}$ and covariance matrix $\boldsymbol{C}_p$ of $\boldsymbol{P}$ is calculated as:

$$\bar{\boldsymbol{p}} = \frac{1}{N_p}\sum_{i=1}^{N_p} \boldsymbol{p}_i, \tag{1}$$

$$\boldsymbol{C}_p = \frac{1}{N_p}\sum_{i=1}^{N_p} \boldsymbol{p}_i\boldsymbol{p}_i^{\mathrm{T}} - \bar{\boldsymbol{p}}\bar{\boldsymbol{p}}^{\mathrm{T}}. \tag{2}$$

Then, an eigenvalue decomposition is performed on the covariance matrix $\boldsymbol{C}_p$ to produce two matrices $\boldsymbol{V}_p$ and $\boldsymbol{E}_p$, where each column in $\boldsymbol{V}_p$ corresponds to an eigenvector of $\boldsymbol{C}_p$, and each diagonal value in the diagonal matrix $\boldsymbol{E}_p$ corresponds to an eigenvalue of $\boldsymbol{C}_p$.

Next, $\boldsymbol{P}$ is aligned with the principal axes defined by $\boldsymbol{V}_p$:

$$\widetilde{\boldsymbol{P}} = \boldsymbol{V}_p\left(\boldsymbol{P} - \bar{\boldsymbol{p}}\right). \tag{3}$$

Consequently, the pose of the 3D face is normalized, resulting in a normalized point cloud $\widetilde{\boldsymbol{P}}$. The corrected 3D face is further resampled with a uniform resolution of 1mm on the xy plane, and the aforementioned pose normalization process is repeated for the resampled point cloud, until the resulting matrix $\boldsymbol{V}_p$ is close to an identity matrix. For simplicity, we use $\boldsymbol{P}$ to denote the normalized point cloud $\widetilde{\boldsymbol{P}}$ in the rest of this paper.

2.2. 3D Keypoint Detection

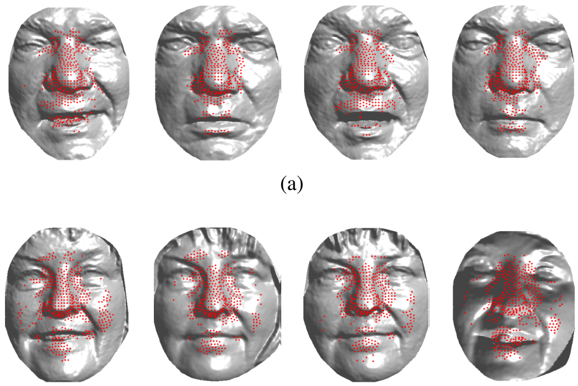
The task of 3D keypoint detection is to select a subset of points with a high discriminative power and a high repeatability from the point cloud of a 3D face [31, 19, 6]. The detected 3D keypoints should be highly robust to noise, pose variations, and various facial expressions. Besides, the feature descriptors extracted from the local surfaces around these keypoints should be sufficiently discriminative for face recognition. In this paper, we use a modified 3D keypoint detection algorithm based on the work in [31].

Given a point cloud $\boldsymbol{P} = \left\{\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_{N_p}\right\} \in \mathbb{R}^3$ after preprocessing, it is first uniformly resampled on the xy plane with a resolution of 4mm, resulting in a set of sample points. For each sample point $\boldsymbol{p}_s$, its neighboring points with distances less than a radius r_k are cropped from the 3D face to form a point set $\boldsymbol{P}_s$. In this paper, the radius r_k is empirically set to 20mm. In order to further select a few highly repeatable keypoints from these sample points in $\boldsymbol{P}_s$, we use the Hotelling transform to calculate the principal axes of the points $\boldsymbol{P}_s$. The point set $\boldsymbol{P}_s$ is then aligned with its principal axes to produce an aligned point set $\widehat{\boldsymbol{P}}_s$. Then, a shape variation index is calculated as the ratio between the surface extensions along the x and y axes:

$$\epsilon = \frac{\max\left(x|x \in \widehat{\boldsymbol{P}}_s\right) - \min\left(x|x \in \widehat{\boldsymbol{P}}_s\right)}{\max\left(y|y \in \widehat{\boldsymbol{P}}_s\right) - \min\left(y|y \in \widehat{\boldsymbol{P}}_s\right)}. \tag{4}$$

This shape variation index reflects the geometric variation of the local surface around a keypoint, and it is different from the one used in [31]. For a symmetric local surface (e.g., a plane or a sphere), the index is 1. For an asymmetric local surface, the index is larger than 1. We consider the sample points with shape variation indices larger than a threshold τ_ϵ as keypoints. The threshold τ_ϵ determines both the repeatability and quantity of keypoints. For a large threshold, the repeatability of keypoints is high, but the number of detected keypoints is small. Therefore, in practice, a tradeoff should be made to select the appropriate threshold (more results and discussions can be found in Section 5.2.1).

In order to better illustrate the proposed keypoint detection algorithm, the keypoints detected on 3D faces with different expressions of two individuals are shown in Fig. 3. It can be observed that most keypoints are detected from areas with large shape variations, including nose and mouth. Although a few keypoints (e.g., those around the mouth) are changed due to facial expressions, the majority of keypoints can still be repeatably detected from the 3D faces of the same individual. Besides, the distribution of keypoints detected from the 3D faces of different individuals varies significantly. For example, most keypoints of the first individual lie around the nose region. However, keypoints can be found in both the nose and cheek regions for the second individual. Note that, the difference in the keypoint distribution among different individuals



(a)

(b)

Fig. 3: 3D keypoint distribution on the faces of two individuals. (a) Keypoints on faces of individual 1. (b) Keypoints on faces of individual 2.

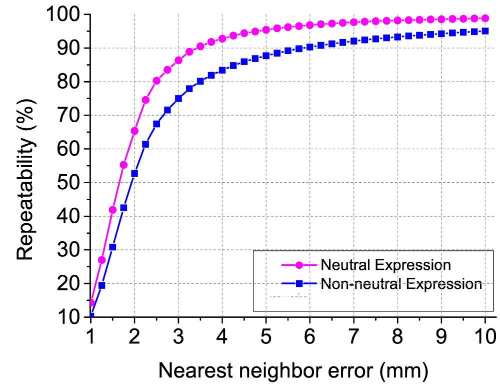


Fig. 4: Repeatability of 3D keypoints.

can be employed to improve the face recognition performance, as demonstrated in Sections 4 and 5.5.

We further analyze the 3D keypoint detection algorithm performance following the same approach as [31]. The experiments were conducted on the Face Recognition Grand Challenge (FRGC) v2 dataset, which contains 4007 3D faces of 466 individuals. Since the ground truth correspondence between 3D faces is unknown, we first align all the faces belonging to the same individual using the Iterative Closest Point (ICP) algorithm [7, 21]. Given a pair of aligned 3D faces, for each keypoint on the first 3D face, we find its closest keypoint on the second 3D face, if their distance is smaller than a threshold τ_k, then the keypoint is considered to be repeatable. The repeatability of 3D keypoints is calculated as the ratio between the number of repeatable keypoints and the total number of keypoints. Finally, the average repeatability of 3D keypoints is calculated over all 3D faces. The repeatability with respect to varying thresholds is shown in Fig. 4. It is clear that, the repeatability drops significantly when τ_k is less than 4mm. That is because the interval between the sample points is 4mm, and it is difficult to detect repeatable keypoints with a distance less than the sampling interval. When τ_k is set to 4mm, the repeatability values achieved on neutral and non-neutral 3D faces are 92.9% and 83.2%, respectively. To better understand the 3D keypoint detection performance, the average number of keypoints detected on 3D faces are presented in Table 1. It is clear that more keypoints can be detected from non-neutral 3D faces than neutral 3D faces. That is reasonable since more regions with significant geometric variations can be found on a non-neutral 3D face. Moreover, although the repeatability of detected keypoints on non-neutral 3D faces is lower than the neutral 3D faces, the number of repeatable keypoints detected on non-neutral 3D faces is comparable to that achieved on neutral 3D faces. The large number of repeatable keypoints ensures that the face recognition algorithm can produce a promising performance, even when tested on 3D faces with large expressions (as shown in Section 5.5).

3. 3D Feature Description and Matching

3.1. 3D Feature Description

Once the keypoints are detected from the 3D facial scans, a feature descriptor is generated from the local surface around each keypoint. In this paper, the Rotational Projection Statistics (RoPS) descriptor [18] is used to encode the geometric information of the corresponding local surface. The RoPS descriptor has been successfully used for 3D object recognition and 3D modeling [18, 20]. In this paper, it is the first time that RoPS is used for 3D face recognition, with promising performance being achieved (Section 5.5).

Given a keypoint $\boldsymbol{q}$ and its support radius r, the neighboring points around keypoint $\boldsymbol{q}$ with distances less than r are cropped from the 3D face, resulting in a point set $\boldsymbol{Q}$. The RoPS descriptor is then generated following procedure below.

First, in order to record the geometric information of $\boldsymbol{Q}$ from different viewpoints, the 3D point set $\boldsymbol{Q}$ is rotated around the x axis by a set of angles $\{\theta_k\}$. For each rotation, the rotated point set $\boldsymbol{Q}'$ is then projected on the xy, yz, and xz coordinate planes, resulting in three 2D point sets $\widetilde{\boldsymbol{Q}'_i}, i = 1, 2, 3$. By projecting the 3D point set onto three 2D planes, the geometric information in $\boldsymbol{Q}$ under that particular viewpoint can be preserved, and the dimensionality is significantly reduced.

Second, for each projected 2D point set $\widetilde{\boldsymbol{Q}'_i}$, its geometric information have to be extracted. For this purpose, the bounding box of $\widetilde{\boldsymbol{Q}'_i}$ is equally divided into $N_b \times N_b$ bins. For each bin, the point number of $\widetilde{\boldsymbol{Q}'_i}$ falling into that bin is counted, resulting in a distribution matrix $\boldsymbol{D}$. The distribution matrix $\boldsymbol{D}$ is further normalized to achieve invariance to point density variations. Since the dimensionality of $\boldsymbol{D}$ is still too high (i.e., $N_b \times N_b$) , the information in $\boldsymbol{D}$ is further encoded with four central moments $\{u_{11}, u_{21}, u_{12}, u_{22}\}$ and the Shannon entropy e. That is:

$$u_{mn} = \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} \left(i - \bar{i}\right) \left(j - \bar{j}\right) \boldsymbol{D}(i, j), \tag{5}$$

$$e = -\sum_{i=1}^{N_b} \sum_{j=1}^{N_b} \boldsymbol{D}(i, j) \log\left(\boldsymbol{D}(i, j)\right). \tag{6}$$

Third, the central moments and Shannon entropy generated from all rotations and projections are concatenated to form a

Table 1: The average number of keypoints detected on 3D faces.

	#Keypoints	#Repeatable Keypoints	Repeatability
Neutral 3D Faces	351	326	92.9%
Non-neutral 3D Faces	368	306	83.2%

sub-feature descriptor for the rotations around the x axis. In order to encode more information of the local surface, $\boldsymbol{Q}$ is also rotated around the y and z axes to generate another two sub-feature descriptors. All these sub-feature descriptors are finally combined to obtain the overall RoPS feature descriptor. Here, the rotation number controls the computational efficiency, the feature dimensionality and the feature descriptiveness. In this paper, the rotation number is set to 3 to achieve a compromise between these considerations. Consequently, the length of our RoPS feature descriptor is $3 \times 3 \times 3 \times 5 = 135$.

Finally, the RoPS feature descriptor is further compressed using PCA [23]. Specifically, a set of training RoPS features are selected and their covariance matrix $\boldsymbol{C}$ is calculated. An eigenvalue decomposition is then applied to $\boldsymbol{C}$ to obtain its eigenvectors. These eigenvectors are rearranged according to the descending order of eigenvalues. The first N_{sf} eigenvectors are used to form a matrix $\boldsymbol{V}_{sf}$. The number N_{sf} is determined such that a ratio ϑ of the fidelity of the training RoPS features is preserved in the compressed features. ϑ is usually a positive number which is close to 1. For a RoPS feature $\boldsymbol{f}_i$, its compressed RoPS feature $\hat{\boldsymbol{f}}_i$ is calculated as

$$\hat{\boldsymbol{f}}_i = \boldsymbol{V}_{sf}^{\mathrm{T}} \boldsymbol{f}_i. \tag{7}$$

3.2. 3D Feature Matching

Assume that $\mathcal{F}^i = \left\{ f_n^i \right\}$ and $\mathcal{F}^j = \left\{ f_m^j \right\}$ are the sets of RoPS features extracted from 3D faces $\boldsymbol{P}^i$ and $\boldsymbol{P}^j$, respectively. The Nearest Neighbor Distance Ratio (NNDR) approach [32] is used to perform feature matching. Specifically, each feature f_n^i in $\mathcal{F}^i$ is matched against all the features in $\mathcal{F}^j$ to obtain its nearest feature $f_{m'}^j$ and the second nearest feature $f_{m''}^j$, that is:

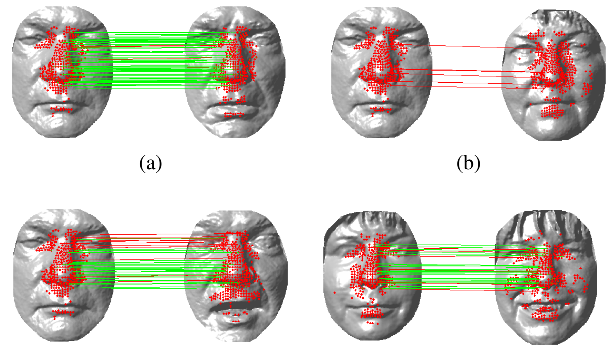
$$f_{m'}^j = \arg \min_{f_m^j \in \mathcal{F}^j} \left\| f_n^i - f_m^j \right\|_2, \tag{8}$$

$$f_{m''}^j = \arg \min_{f_m^j \in \mathcal{F}^j \setminus f_{m'}^j} \left\| f_n^i - f_m^j \right\|_2, \tag{9}$$

where $\mathcal{F}^j \setminus f_{m'}^j$ is the feature set $\mathcal{F}^j$ excluding feature $f_{m'}^j$. NNDR r_{dis} is calculated as:

$$r_{dis} = \frac{\left\| f_n^i - f_{m'}^j \right\|_2}{\left\| f_n^i - f_{m''}^j \right\|_2}. \tag{10}$$

If the ratio r_{dis} is less than a threshold τ_f, $\left(f_n^i, f_{m'}^j \right)$ is considered as a potential feature match. To achieve robust feature matching, $f_{m'}^j$ is also matched against all the features in $\mathcal{F}^i$. If f_n^i is the nearest feature in $\mathcal{F}^i$ for $f_{m'}^j$ and satisfies the NNDR criterion, then $\left(f_n^i, f_{m'}^j \right)$ is finally confirmed as a feature match. The threshold τ_f controls both the number and accuracy of feature matches. A small threshold produces a limited number of feature matches and is not sufficient for all accurate transformation estimation. In contrast, a large threshold results in a large number of false positive matches, which also degrades the performance of the transformation estimation. The face recognition performance with different thresholds is further analyzed in Section 5.2.3. Figure 5 presents an illustration of the feature matching for 3D faces of an individual with a neutral expression (Fig. 5(a)), 3D faces of two individuals with a neutral expression (Fig. 5(b)), 3D faces of an individual with different expressions (Fig. 5(c)), and 3D faces of an individual with different expressions and hair occlusions (Fig. 5(d)). If the spatial distance between two matched features is less than 4mm, the feature match is considered correct, and is denoted by the green lines in Fig. 5. Otherwise, the feature match is considered false, and is denoted by the red lines in Fig. 5. It can be seen that a number of features are correctly matched for two 3D faces from the same individual, even with different facial expressions and hair interference. However, the majority of features from two different individuals cannot be correctly matched, even with a neutral facial expression. Consequently, the feature matching algorithm can be used to identify the same individual with different expressions, and to distinguish different individuals with the same expression.



(a) (b) (c) (d)

Fig. 5: An illustration of the feature matching results. (a) Faces of an individual with a neutral expression. #feature matches: 85, #false matches: 4; (b) Faces of two individuals with a neutral expression. #feature matches: 6, #false matches: 6; (c) Faces of an individual with different expressions. #feature matches: 42, #false matches: 13; (d) Faces of an individual with different expressions and hair occlusions. #feature matches: 41, #false matches: 8. (Figure best seen in color.)

We match all of the features in $\mathcal{F}^i$ against the features in $\mathcal{F}^j$, resulting in a set of matched keypoints $C^{ij} = \left\{ c_1^{ij}, c_2^{ij}, \ldots, c_{N_c}^{ij} \right\}$, where $c_n^{ij} = \left\{ q_n^i, q_n^j \right\}$ is a pair of matched keypoints. It is sensible to use the feature matching results to measure the similarity between two 3D faces. In this paper, both the number of feature matches n_{fc} and the average distance of matched features d_{fc} are used as two metrics for 3D face similarity calculation. Further, we used the Procrustes algorithm [16] to perform registration between the two 3D faces $\boldsymbol{P}^i$ and $\boldsymbol{P}^j$. The number of closest point pairs with distances less than 2mm are calculated, as denoted by n_{pc}. Finally, n_{fc}, d_{fc} and n_{pc} are considered as three metrics to measure the similarity between 3D faces. Note that, the major task of this paper is to present a general framework for automatic 3D face recognition, other feature similarity metrics can also be integrated into the proposed framework, including the graph matching approach [31] and the RANSAC algorithm [5, 6]. In order to improve the computational efficiency, we only consider the aforementioned three metrics in this paper.

4. 3D Face Recognition

Face recognition includes two different tasks: face identification and face verification [3]. For face identification, the probe face is compared with all the gallery faces to obtain the identity of the probe face. Face identification has to calculate the similarities between the probe face and all gallery faces, it is therefore a one-vs-all matching process. For face verification, the probe face is compared with the claimed face to determine whether the two faces belong to the same person. Face verification has to calculate only the similarity between the probe face and the claimed face, it is therefore a one-vs-one matching process.

During offline processing, a gallery with N_g neutral faces is constructed, with each face corresponding to an individual. The 3D keypoints and compressed RoPS features are extracted from each face and stored in the gallery. During online recognition, the keypoints and compressed RoPS features are first extracted. For face identification, the probe face is matched against all the gallery faces using the RoPS feature matching and point cloud registration, resulting in three similarity metrics n_{fc}, d_{fc}, n_{pc}. For each metric, a similarity vector $\boldsymbol{s}_k$ can be obtained ($k = 1, 2, 3$). The m-th element s_{km} in $\boldsymbol{s}_k$ represents the similarity between the probe face and the m-th gallery face using the k-th similarity metric. In order to further improve the face recognition performance, the similarity results achieved by these metrics are fused. For unbiased fusion, each similarity $\boldsymbol{s}_k$ have to be normalized to the range of [0,1] using the min-max rule. That is,

$$\widehat{\boldsymbol{s}}_k = \frac{\boldsymbol{s}_k - \min(\boldsymbol{s}_k)}{\max(\boldsymbol{s}_k - \min(\boldsymbol{s}_k)) - \min(\boldsymbol{s}_k - \min(\boldsymbol{s}_k))}. \tag{11}$$

$\widehat{\boldsymbol{s}}_1$ and $\widehat{\boldsymbol{s}}_3$ have a positive polarity, i.e., a large value of $\widehat{\boldsymbol{s}}_k$ represents a high similarity. Since $\widehat{\boldsymbol{s}}_2$ has a negative polarity, it is further normalized to achieve a positive polarity, that is:

$$\widehat{\boldsymbol{s}}_2 = 1 - \widehat{\boldsymbol{s}}_2. \tag{12}$$

Once these similarities have been calculated, a fused similarity is calculated as:

$$\boldsymbol{s} = \sum_{k=1}^{3} \omega_k \widehat{\boldsymbol{s}}_k. \tag{13}$$

where ω_k is the weight for the k-th similarity metric, which can be learned from the training stage (see Section 5.4).

The fused similarity is further normalized as:

$$\widehat{\boldsymbol{s}} = \frac{\boldsymbol{s} - \min(\boldsymbol{s})}{\max(\boldsymbol{s} - \min(\boldsymbol{s})) - \min(\boldsymbol{s} - \min(\boldsymbol{s}))}. \tag{14}$$

For face identification, the identity of the probe face is determined by the gallery face with the highest similarity. For face verification, the probe face is considered to be from the claimed individual if the similarity is above a set threshold.

Note that, our method uses local geometric features rather than holistic features for 3D face recognition. Consequently, the proposed method can cope with various facial expressions. This advantage of our method is due to several factors. **First**, for a 3D face with facial expressions, since a large number of keypoints have been extracted from the 3D face, the keypoints extracted from the unaffected areas (without occlusions) are still sufficient for feature matching and face registration. Therefore, the face with facial expressions can still be correctly recognized. **Second**, even if a part of the facial shape is deformed by facial expressions, the features from the occluded part of the 3D face cannot be matched with the features from the gallery face. Only the features from the unaffected areas of the 3D face can contribute to the feature matching results. Therefore, facial expressions can be well handled by our method. **Third**, the shape deformation of the rigid and semi-rigid regions (e.g., nose and forehead) of a face is small under different facial expressions, while the deformation of the other regions (e.g., month and cheek) is relatively large. For a face with large expressions, the features extracted from the rigid/semi-rigid areas can still achieve a correct feature matching with the gallery faces. Therefore, the effect of facial expressions can significantly be reduced by our algorithm.

5. Experimental Results

5.1. Experimental Setup

5.1.1. Dataset Description

In this paper, we use the publically available FRGC dataset [37] to test our proposed EI3D algorithm. The dataset includes 4950 3D facial scans with shoulder level up from 466 individuals. These facial scans were acquired using a Minolta Vivid 900/910 scanner and are divided into three subsets, i.e., Spring2003, Fall2003, and Spring2004. The dataset is further partitioned into a training dataset (FRGC v1) and a validation dataset (FRGC v2). The training dataset (FRGC v1) includes 943 3D facial scans from the Spring2003 subset, while the validation dataset (FRGC v2) includes 4007 3D facial scans from the Fall2003 and Spring2004 datasets. The validation dataset contains 2410 facial scans with neutral expression, and 1597 facial scans with various facial expressions including disgust, happiness, sadness, surprise, and anger [6, 24]. Besides, other nuisances can be found in the 3D facial scans, including noise, spikes, holes, and hair occlusions.

In this paper, the first neutral facial scan was selected from the scans of each individual to form a gallery of 466 facial scans. The remaining 3541 facial scans were used to form the test dataset. Consequently, the test dataset consists of 1944 neutral facial scans and 1597 non-neutral facial scans.

Table 2: The keypoint number and face verification performance under different thresholds of the keypoint detection.

Threshold τ_ϵ	1.04	1.06	1.08	1.10	1.12
#keypoints	360	241	170	125	97
VR@0.1% FAR (%)	98.33	98.07	97.58	96.31	93.27

Table 3: Face verification performance under different feature matching thresholds.

Threshold	0.7	0.8	0.9	1.0
VR@0.1%FAR	98.47	98.22	98.22	97.2

5.1.2. Evaluation Criteria

For face identification, we adopt the frequently used Cumulative Match Characteristic (CMC) and Rank-1 Recognition Rate (R1RR) to measure the performance. The CMC curve presents the percentage of correctly recognized probe faces with respect to the rank number that is considered as a correct recognition, while R1RR is the percentage of the probe faces that are correctly recognized using the first rank. For face verification, we use the Receiver Operating Characteristic (ROC) and the verification rate (VR) at a false acceptance rate (FAR) of 0.1% (VR@0.1%FAR) to measure the performance. The horizontal axis of the ROC curve is the False Accept Rate (FAR) while the vertical axis is the Verification Rate (VR). For more details on these evaluation criteria, the reader should refer to [3].

5.2. Performance with Different Parameters

In this section, we tested our algorithm with respect to different settings of the threshold τ_ϵ for keypoint detection, the support radius r for feature description and the threshold τ_f for feature matching. The face verification results on the FRGC v1 dataset under different parameter settings are shown in Fig. 6.

5.2.1. The Threshold for Keypoint Detection

The threshold τ_ϵ determines both the number and the repeatability of the detected keypoints. We tested our face verification algorithm with a set of threshold values for τ_ϵ, namely 1.04, 1.06, 1.08, 1.10 and 1.12. The support radius r is set to 15mm, the threshold τ_f is set to 0.7, and no feature compression is applied in this experiment. The ROC results are shown in Fig. 6(a). It is clear that the face verification performance drops as the threshold is increased. When τ_ϵ is set to 1.04, the VR@0.1%FAR value is 98.33%. Then, when τ_ϵ is set to 1.08, the VR@0.1%FAR value is 97.58%. When τ_ϵ is further increased to 1.12, the VR@0.1%FAR value is only 93.27%. It can also be observed from Table 2 that, as the threshold τ_ϵ increases, the number of keypoints decreases, resulting in a low feature matching accuracy. Note that, more keypoints is usually beneficial for the improvement of the feature matching, especially for faces with expressions. Based on these analyses, τ_ϵ was set to 1.04 in this paper.

5.2.2. The Support Radius for Feature Description

The support radius determines both the discriminative power and the robustness with respect to expressions. We tested our face verification algorithm with a support radius set to 5mm, 10mm, 15mm, 20mm and 25mm. The threshold τ_f was set to 0.7, and no feature compression was applied in this experiment. The ROC results are shown in Fig. 6(b). It can be seen that the face verification performance increases significantly as the support radius is increased from 5mm to 10mm. That is because when the support radius is small, the discriminative power of the feature descriptor is insufficient. When the support radius is further increased from 10mm to 15mm, the method achieves its best performance. When the support radius is further increased, the face verification performance is decreased. That is because a trade-off has been made between the discriminative power and the robustness when the support radius is set to 15mm. A large support radius makes the extracted feature descriptor sensitive to expressions, therefore, the overall verification performance is degraded. In this paper, the support radius is set to 15mm.

5.2.3. The Threshold for Feature Matching

The threshold τ_f determines both the number and accuracy of matched features. A small τ_f results in a high accuracy of feature matching, but the number of matched features is small. In this section, we tested the face verification performance with a threshold τ_f set to 0.7, 0.8, 0.9 and 1.0, and the results are shown in Table. 3. It is clear that the best performance is achieved when τ_f is set to 0.7. When the threshold is increased, the recognition performance decreases slightly. That is because many false feature matches are encountered when the threshold is large, and therefore the recognition performance is decreased by these false matches. In this paper, the threshold τ_f is set to 0.7 for the subsequent experiments.

5.3. Performance with Feature Compression

The PCA preservation rate ϑ determines both the length of the compressed feature and the preserved information after compression. In this section, we tested our face verification algorithm on the FRGC v1 dataset with a preservation rate set to 86%, 88%, 90%, 92%, 94%, 96%, 98% and 100%. The threshold τ_f for feature matching was set to 0.7. The ROC results are shown in Fig. 6(c). It can be observed that the face verification performance increases steadily as the rate ϑ is increased from 86% to 98%. The performance is then decreased as the percentage ϑ is further increased above 98%. That is because for the rate 98%, the useful discriminative information in the features have been preserved, and the undesired information caused by expression variations is discarded. Consequently, the best performance can be achieved. It can also be noticed that the performance achieved by uncompressed features is even inferior to the compressed features with a preservation rate of 98%. This observation clearly shows that feature compression can reduce the unnecessary information contained in the features and improve the accuracy of feature matching. From Table 4, it can be seen that the feature length is reduced from 135 to 30 when the rate ϑ is set to 98%, and the VR@0.1%FAR is as high as 98.47%. Based on the above analyses, the preservation rate ϑ is set to 98% in this paper.

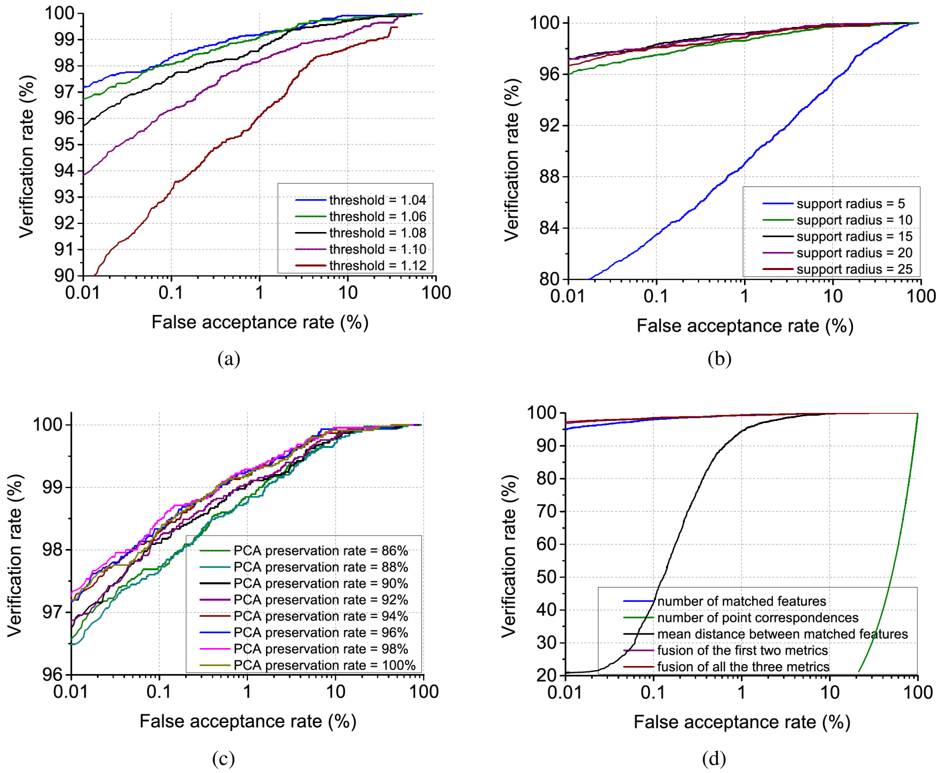


Fig. 6: Face verification results on the FRGC v1 dataset under different parameter settings. (a) Results with different thresholds of keypoint detection. (b) Results with different support radii for feature description. (c) Results with different preservation rates for feature compression. (d) Results with different similarity metrics. (Figure best seen in color.)

Table 4: The feature length and face verification performance under different fidelity percentages.

Percentage ϑ	86	88	90	92	94	96	98	100
Feature length	12	13	14	16	19	22	30	135
VR@0.1% FAR (%)	97.73	97.67	98.11	98.20	98.29	98.33	98.47	98.33

5.4. *Performance with Metric Fusion*

In this section, we tested the face verification performance on the FRGC v1 dataset using different similarity metrics and their combinations, the results are reported in Fig. 6(d). It is clear that metric n_{fc} achieves the best performance with a VR@0.1%FAR of 98.05%, followed by n_{pc} with a VR@0.1%FAR of 96.42%. In contrast, metric d_{fc} achieves the worst performance, with a VR@0.1%FAR of only 42.17%. We further tested the face verification system with a similarity fusion (as defined in Eq. 13), the weight for each metric is determined by its VR@0.1%FAR value using a single similarity metric. It can be seen that the recognition performance using both n_{fc} and n_{pc} is the same as the one achieved using all these three metrics, with a VR@0.1%FAR of 98.47%. It can be inferred that fusing the information of n_{fc} and n_{pc} is beneficial to the improvement of face recognition performance. However, adding d_{fc} cannot improve the 3D face recognition performance since the performance achieved by metric d_{fc} is very low. Consequently, we use the fusion results of n_{fc} and n_{pc} for face recognition in the rest of this paper.

5.5. *Comparison with the State-of-the-Art*

The FRGC v2 dataset was used to test the face verification and identification performance of our proposed algorithm.

5.5.1. *Face Verification*

The 3D face verification results on the FRGC v2 dataset is shown in Fig. 7 (b). It can be seen that the VR@0.1%FAR achieved by our algorithm is as high as 99.9% for neutral faces. This means that our algorithm is highly suitable for face verification applications. Even for non-neutral faces, our algorithm still achieves a high VR@0.1%FAR of 97.12%. This means that our algorithm can cope with facial expressions and can perform accurate face verification for non-cooperative individuals. Our algorithm achieves an average VR@0.1%FAR of 99.01% on all faces with neutral and non-neutral faces.

To compare our results with the state-of-the-art results achieved on the FRGC v2 dataset, we present the VR@0.1%FAR results of existing algorithms in Table 5. It can be seen that our algorithm achieves the best face verification performance on neutral 3D faces, with a VR@0.1%FAR of 99.9%. For non-neutral 3D faces, our algorithm also achieves a VR@0.1%FAR of 97.18%, which is close to the best results reported in the literature (e.g., 97.8%). For the entire dataset containing both neutral and non-neutral 3D faces, our algorithm is superior to all the existing algorithms, achieving a high verification rate of 99.01%.

5.5.2. Face Identification

The 3D face identification results on the FRGC v2 dataset is shown in Fig. 7 (a). It can be seen that our algorithm achieved a Rank-1 recognition rate of 99.4% for neutral faces, which means that our algorithm can be used to accurately recognize 3D faces under neutral expressions. For 3D faces with various expressions including disgust, happiness, sadness, surprise, and anger, the performance of our algorithm decreases slightly. However, its rank-1 recognition rate is still as high as 94.0%. This clearly demonstrates that our algorithm is robust to non-rigid deformations caused by facial expressions. The overall Rank-1 recognition rate for all 3D faces is 97.0%.

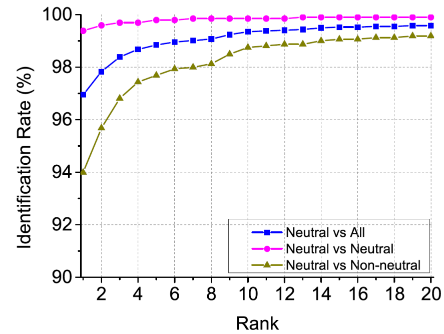
The high recognition rate and strong robustness of our 3D face recognition algorithm is due to several facts. **First**, the 3D keypoints detected by our algorithm has a high repeatability (see Section 2.2). Although facial deformation will introduce variations in the locations of the 3D keypoints, the majority of the 3D keypoints can still be robustly detected. Therefore, the final 3D face recognition performance is insensitive to facial expressions. **Second**, the extracted RoPS local feature descriptor is highly discriminative and highly robust to nuisances including facial expressions, which ensures the high accuracy and strong robustness achieved by our proposed 3D face recognition algorithm. **Third**, the face recognition performance is further boosted through the fusion between two different similarity metrics.

6. Conclusion

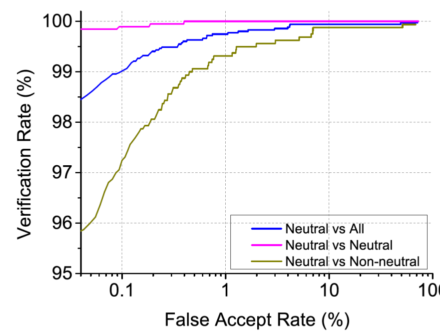
In this paper, we propose an accurate expression-invariant face recognition algorithm based on local feature matching and shape registration. A 3D face is represented by a set of class-specific keypoints, and then described with their associated RoPS local features. Face similarity is calculated using the feature matching and shape registration metrics to produce face identification and verification results. The proposed algorithm not only fully employs the global similarity information between faces using face registration, but also inherits the strong robustness brought by the local features. Experimental results on the FRGC dataset show that the proposed algorithm achieves high face identification and verification rates. Moreover, our algorithm is robust to expression variations.

Acknowledgments

This research is supported by three National Natural Science Foundation of China grants (No. 61471371, No. 61403265 and No. 61202336). This work is also partly supported by the Science and Technology Plan of Sichuan Province (No. 2015SZ0226).



(a)



(b)

Fig. 7: Face recognition results on the FRGC v2 dataset. (a) CMC results for face identification. (b) ROC results for face verification. (Figure best seen in color.)

References

[1] F. Al-Osaimi, M. Bennamoun, and A. Mian. Integration of local and global geometrical cues for 3D face recognition. *Pattern Recognition*, 41 (3):1030–1040, 2008.

[2] F. Al-Osaimi, M. Bennamoun, and A. Mian. An expression deformation approach to non-rigid 3D face recognition. *International Journal of Computer Vision*, 81(3):302–316, 2009.

[3] M. Bennamoun, Y. Guo, and F. Sohel. Feature selection for 2D and 3D face recognition, in: Encyclopedia of electrical and electronics engineering. *Book Chapter*, pages 1–54, 2015.

[4] S. Berretti, A. Del Bimbo, and P. Pala. 3D face recognition using isogeodesic stripes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2162–2177, 2010.

[5] S. Berretti, N. Werghi, A. del Bimbo, and P. Pala. Matching 3D face scans using interest points and local histogram descriptors. *Computers & Graphics*, 2013.

[6] S. Berretti, N. Werghi, A. del Bimbo, and P. Pala. Selecting stable keypoints and local descriptors for person identification using 3D face scans. *The Visual Computer*, pages 1–18, 2014.

[7] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2): 239–256, 1992.

[8] C. Beumier and M. Acheroy. Face verification from 3D and grey level clues. *Pattern Recognition Letters*, 22(12):1321–1329, 2001.

[9] K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Computer vision and image understanding*, 101(1):1–15, 2006.

[10] Kevin W Bowyer, Kyong I Chang, and Patrick J Flynn. A survey of approaches to three-dimensional face recognition. In *International Conference on Pattern Recognition*, pages 358–361, 2004.

[11] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Expression-invariant representations of faces. *IEEE Transactions on Image Processing*, 16(1): 188–197, 2007.

[12] C. S. Chua and R. Jarvis. Point signatures: a new representation for 3D object recognition. *International Journal of Computer Vision*, 25(1):63–85, 1997.

Table 5: Comparison with state-of-the-art approaches.

	Neutral vs All	Neutral vs Neutral	Neutral vs Non-neutral
Mian et al. [30], 2007	98.5%	99.4%	97.0%
Mian et al. [31], 2008	97.4%	99.9%	92.7%
Al-Osaimi et al. [2], 2009	98.14%	98.35%	**97.8%**
Berretti et al. [4], 2010	95.5%	97.7%	91.4%
Wang et al. [44], 2010	98.61%	99.2%	97.7%
Lei et al. [25], 2014	NA	NA	97.8%
This paper	**99.01%**	**99.9%**	97.18%

[13] C. S. Chua, F. Han, and Y. K. Ho. 3D human face recognition using point signature. In *4th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 233–238, 2000.

[14] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama. 3D face recognition under expressions, occlusions and pose variations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.

[15] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[16] C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 285–339, 1991.

[17] G. Gordon. Face recognition based on depth and curvature features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 808–810, 1992.

[18] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan. Rotational projection statistics for 3D local surface description and object recognition. *International Journal of Computer Vision*, 105(1):63–86, 2013.

[19] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan. 3D object recognition in cluttered scenes with local surface features: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2270–2287, 2014.

[20] Y. Guo, F. Sohel, M. Bennamoun, J. Wan, and M. Lu. An accurate and robust range image registration algorithm for 3D object modeling. *IEEE Transactions on Multimedia*, 16(5):1377–1390, 2014.

[21] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan. An integrated framework for 3D modeling, object detection and pose estimation from point-clouds. *IEEE Transactions on Instrumentation and Measurement*, 64(3):683–693, 2015.

[22] S. Gupta, M. K. Markey, and A. C. Bovik. Anthropometric 3D face recognition. *International Journal of Computer Vision*, 90(3):331–349, 2010.

[23] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 498–506, 2004.

[24] Y. Lei, M. Bennamoun, and A.A. El-Sallam. An efficient 3D face recognition approach based on the fusion of novel local low-level features. *Pattern Recognition*, 46(1):24–37, 2013.

[25] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo. An efficient 3D face recognition approach using local geometrical signatures. *Pattern Recognition*, 47(2):509–524, 2014.

[26] L. Li, C. Xu, W. Tang, and C. Zhong. 3D face recognition by constructing deformation invariant image. *Pattern Recognition Letters*, 29(10):1596–1602, 2008.

[27] L. Liu and P. W. Fieguth. Texture classification from random features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):574–586, 2012.

[28] L. Liu, P. Fieguth, G. Zhao, and M. Pietikainen. Extended local binary pattern fusion for face recognition. In *IEEE International Conference on Image Processing*, pages 718–722. IEEE, 2014.

[29] P. Liu, Y. Wang, D. Huang, Z. Zhang, and L. Chen. Learning the spherical harmonic features for 3-D face recognition. *IEEE Transactions on Image Processing*, 22(3):914–925, 2013.

[30] A.S. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11):1927–1943, 2007.

[31] A.S. Mian, M. Bennamoun, and R. Owens. Keypoint detection and local feature matching for textured 3D face recognition. *International Journal of Computer Vision*, 79(1):1–12, 2008.

[32] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[33] H. Mohammadzade and D. Hatzinakos. Iterative closest normal point for 3D face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):381–397, 2013.

[34] A. Moreno, A. Sánchez, J. F. Vélez, and F. J. Díaz. Face recognition using 3D surface-extracted descriptors. In *Irish Machine Vision and Image Processing Conference*, volume 2003, 2003.

[35] Maja Pantic and Leon J. M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.

[36] G. Passalis, P. Perakis, T. Theoharis, and I. A. Kakadiaris. Using facial symmetry to handle pose variations in real-world 3D face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):1938–1951, 2011.

[37] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE conference on computer vision and pattern recognition*, volume 1, pages 947–954, 2005.

[38] C. Samir, A. Srivastava, M. Daoudi, and E. Klassen. An intrinsic framework for analysis of facial surfaces. *International Journal of Computer Vision*, 82(1):80–95, 2009.

[39] D. Smeets, P. Claes, D. Vandermeulen, and J. G. Clement. Objective 3D face recognition: Evolution, approaches and challenges. *Forensic science international*, 201(1):125–132, 2010.

[40] D. Smeets, P. Claes, J. Hermans, D. Vandermeulen, and P. Suetens. A comparative study of 3-D face recognition under expression variations. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 42(5):710–727, 2012.

[41] D. Smeets, J. Keustermans, D. Vandermeulen, and P. Suetens. meshSIFT: local surface features for 3D face recognition under expression variations and partial data. *Computer Vision and Image Understanding*, 117(2):158–169, 2013.

[42] A. Srivastava, C. Samir, S. H. Joshi, and M. Daoudi. Elastic shape models for face analysis using curvilinear coordinates. *Journal of Mathematical Imaging and Vision*, 33(2):253–265, 2009.

[43] C. Sun and J. Sherrah. 3D symmetry detection using the extended Gaussian image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):164–168, 1997.

[44] Y. Wang, J. Liu, and X. Tang. Robust 3D face recognition by local shape difference boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10):1858–1870, 2010.

[45] H.-S. Wong, K. Cheung, and H. Ip. 3D head model classification by evolutionary optimization of the Extended Gaussian Image representation. *Pattern Recognition*, 37(12):2307–2322, 2004.

[46] C. Xu, Y. Wang, T. Tan, and L. Quan. Automatic 3D face recognition combining global geometric features with local shape variation information. In *6th IEEE international conference on Automatic face and gesture recognition*, pages 308–313, 2004.

[47] A. Zaharescu, E. Boyer, and R. Horaud. Keypoints and local descriptors of scalar functions on 2D manifolds. *International Journal of Computer Vision*, 100:78–98, 2012.

[48] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *Acm Computing Surveys*, 35(4):399–458, 2003.