



El3D: 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, 410073, Hunan, China

^b College of Electronics and Information Engineering, Sichuan University, Chengdu, 610065, Sichuan, China

^c College of Information System and Management, National University of Defense Technology, Changsha, 410073, Hunan, China

^d College of Computer Science, Sichuan University, Chengdu, Sichuan, China

^e School of Computer Science and Software Engineering, The University of Western Australia, Perth 6009, WA, Australia

^f School of Engineering and Information Technology, Murdoch University, Perth 6150, WA, Australia

ARTICLE INFO

Article history:

Available online 23 April 2016

Keywords:

3D face recognition
Keypoint detection
Shape matching
Facial expression
Local feature
Face identification

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 identification rate of 97.0% and an average verification rate of 99.01% at 0.001 false acceptance rate for all faces with neutral and non-neutral expressions. Further, the robustness of our algorithm under different occlusions has been demonstrated on the Bosphorus dataset.

© 2016 Elsevier B.V. All rights reserved.

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 [20,26]. Facial biometrics has a number of applications including surveillance, security, entertainment, commerce, and forensics [4]. It is particularly suitable for applications where other biometrics (including iris images, retinal scans and fingerprints) are not available or desirable [25].

Face recognition can be performed using 2D facial images, 3D facial scans or their combination [4,19]. 2D face recognition has been extensively investigated during the past few decades. 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 [6,7]. With the rapid

development of 3D scanners, 3D data acquisition is becoming increasingly cheaper and non-intrusive [1]. Besides, 3D facial scans are more robust to lighting conditions, pose variations and facial makeup [7]. 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, and has been considered as an alternative or complementary solution to conventional 2D face recognition approaches [7,20].

Existing 3D face recognition algorithms can be categorized into two classes: holistic and local feature based algorithms [4,21]. 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 [25,27], extended Gaussian images [36], canonical forms [8] and spherical harmonic features [24]. 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 [26].

In contrast, local feature based algorithms detect and match a set of local features (e.g., landmarks, curves, patches) to perform

[☆] "This paper has been recommended for acceptance by Xiang Bai".

* Corresponding author.

E-mail address: yinjie@scu.edu.cn (Y. Lei).

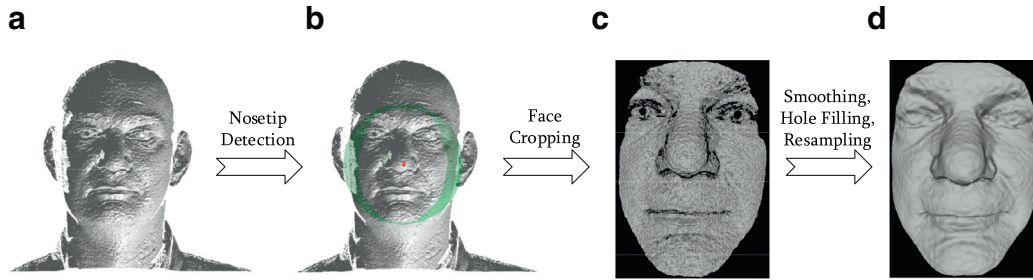


Fig. 1. An illustration of 3D facial data preprocessing. (a) Raw facial scan. (b) Facial scan with detected nosetip. (c) Cropped 3D face. (d) 3D face after spike removal, hole filling and smoothing.

3D face recognition. They are therefore more robust to facial expressions, missing data and occlusion. Local feature based algorithms can further be divided into landmark-based, curve-based, and patch-based algorithms [4]. 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 [4,17]. The curve-based algorithms represent each face with a set of curves, including iso-depth curves, iso-radius curves, iso-geodesic curves and profiles [4,5,33]. 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 [33]. 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 [9], extended Local Binary Patterns (eLBP) [18], Mesh Scale-Invariant Feature Transform (Mesh-SIFT) [34], Mesh Histogram of Oriented Gradients (Mesh-HOG) [6], and local normal patterns [22]. Recently, Li et al. [23] used two principal curvature for 3D keypoint detection and three surface differential quantities for keypoint description. 3D face recognition were finally achieved using a multitask sparse representation. The algorithm achieves the state-of-the-art performance on the Bosphorus and Face Recognition Great Challenge (FRGC) v2 datasets. For more discussions on existing 3D face recognition algorithms, the reader should refer to the recent reviews [4,33].

For a face recognition system, high recognition accuracy and a strong robustness are the two key considerations for many practical applications [1]. 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 variations, the shape of some particular 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 including facial expressions and occlusion [14,26].

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, 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 first tested on the 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 EI3D algorithm was then tested on the Bosphorus dataset. It achieved promising results on faces with different occlusions.

Note that, although the RoPS feature has been successfully used for 3D object recognition [15], its performance on 3D face recognition was still unknown. The task of 3D face recognition is different from 3D generic object recognition in many aspects. First, the 3D facial scans acquired from different individuals can be highly similar while the 3D facial scans acquired from the same individual under different expressions may vary significantly. This introduces great challenges for classification due to the high intra-class difference and small inter-class difference. Second, the major challenges faced by face recognition are expression variation, occlusion, and pose changes. In contrast, the major challenges faced by object recognition are occlusion, clutter, noise and varying mesh resolutions. Consequently, the two tasks are differently defined and require specific designed algorithms. In this paper, we use the RoPS feature for face representation and a well-defined face recognition framework [26] for classification. The performance of our algorithm has been fully demonstrated on the FRGC and the Bosphorus datasets.

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.

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 [20]. Therefore, the raw facial scan have to be preprocessed before any further operations [4,11,20,25,28]. In this paper, we use the work proposed in [25] 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 (as shown in Fig. 1).

2.1.1. Nosetip detection and face cropping

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)) [25]. 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

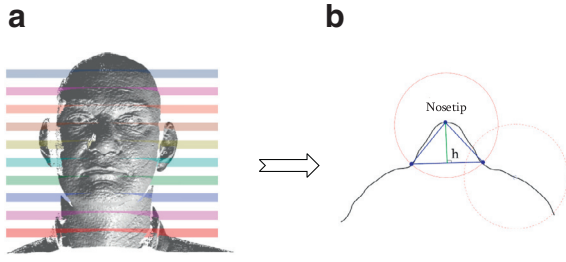


Fig. 2. An illustration of 3D facial nosetip detection. (a) Horizontal planes for 3D facial scan slicing. (b) Horizontal facial profile.

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. The remaining candidates can then be considered lying on the noise ridge and the one with the largest altitude is considered to be the nosetip. Once the nosetip is detected, a 3D face is then cropped from the facial scan by eliminating the points which are located more than 80 mm from the nosetip (as shown in Fig. 1(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 1 mm. However, the spike removal process will result in undesired holes on the 3D face. Besides, the holes can also be caused by other factors including 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 [4]. For 3D faces, these holes can be filled using cubic interpolation. Finally, noise is further smoothed using a median filter (as shown in Fig. 1(d)).

2.1.3. Pose normalization and resampling

Given the point cloud $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{N_p}\} \in \mathbb{R}^3$ of a 3D face, where N_p is the number of points on the 3D face, the Hotelling transform is used to perform pose normalization [25]. First, the mean $\bar{\mathbf{p}}$ and covariance matrix \mathbf{C}_p of \mathbf{P} is calculated as:

$$\bar{\mathbf{p}} = \frac{1}{N_p} \sum_{i=1}^{N_p} \mathbf{p}_i, \quad (1)$$

$$\mathbf{C}_p = \frac{1}{N_p} \sum_{i=1}^{N_p} \mathbf{p}_i \mathbf{p}_i^T - \bar{\mathbf{p}} \bar{\mathbf{p}}^T. \quad (2)$$

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

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

$$\tilde{\mathbf{P}} = \mathbf{V}_p(\mathbf{P} - \bar{\mathbf{p}}). \quad (3)$$

Consequently, the pose of the 3D face is normalized, resulting in a normalized point cloud $\tilde{\mathbf{P}}$. 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 \mathbf{V}_p is close to an

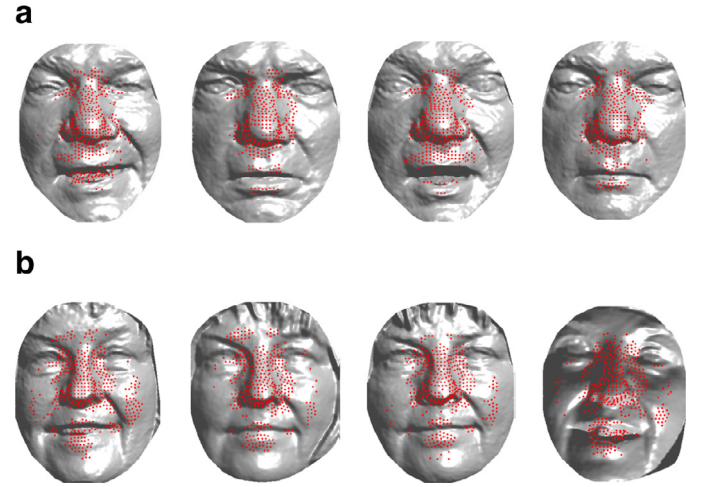


Fig. 3. 3D keypoint distribution on the faces of two individuals. (a) Keypoints on faces of individual 1. (b) Keypoints on faces of individual 2.

identity matrix. For simplicity, we use \mathbf{P} to denote the normalized point cloud $\tilde{\mathbf{P}}$ 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 [7,14,26]. 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 3D keypoint detection algorithm similar to [26].

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

$$\epsilon = \frac{\max(x|x \in \hat{\mathbf{P}}_s) - \min(x|x \in \hat{\mathbf{P}}_s)}{\max(y|y \in \hat{\mathbf{P}}_s) - \min(y|y \in \hat{\mathbf{P}}_s)}. \quad (4)$$

This shape variation index reflects the geometric variation of the local surface around a keypoint, and it is different from the one used in [26]. 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.,

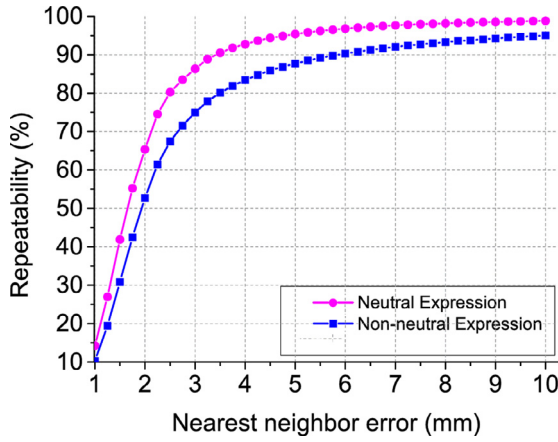


Fig. 4. Repeatability of 3D keypoints.

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 |

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 can be employed to improve the face recognition performance, as demonstrated in Section 4 and 5.

We further analyze the 3D keypoint detection algorithm performance following the same approach as [26]. 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. 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 4 mm. That is because the interval between the sample points is 4 mm, and it is difficult to detect repeatable keypoints with a distance less than the sampling interval. When τ_k is set to 4 mm, 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 recog-

nition algorithm can produce a promising performance, even when tested on 3D faces with large expressions (as shown in Section 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 [15] 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 [15,16]. In this paper, it is the first time that RoPS is used for 3D face recognition, with promising performance being achieved (Section 5).

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

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

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

$$u_{mn} = \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} (i - \bar{i})(j - \bar{j}) \mathbf{D}(i, j), \quad (5)$$

$$e = - \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} \mathbf{D}(i, j) \log(\mathbf{D}(i, j)). \quad (6)$$

Third, the central moments and Shannon entropy generated from all rotations and projections are concatenated to form a sub-feature descriptor for the rotations around the x axis. In order to encode more information of the local surface, \mathbf{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 determines the computational efficiency, the feature dimensionality and the feature descriptiveness. In this paper, the rotation number is set to 3 to achieve a compromise performance 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 the Principal Component Analysis (PCA) technique. Specifically, a set of training RoPS features are selected and their covariance matrix \mathbf{C} is calculated. An eigenvalue decomposition is then applied to \mathbf{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 \mathbf{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. Usually, θ is a positive number close to 1. For a RoPS feature \mathbf{f}_i , its compressed

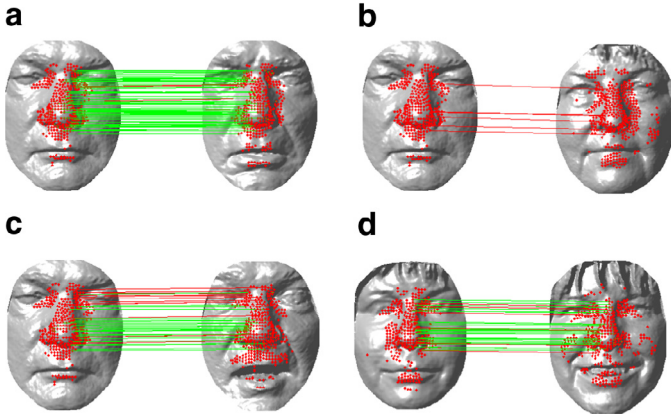


Fig. 5. An illustration of the feature matching results. (a) Faces of an individual with a neutral expression. #feature matches: 85, #false matches: 4; (b) faces of two individuals with a neutral expression. #feature matches: 6, #false matches: 6; (c) faces of an individual with different expressions. #feature matches: 42, #false matches: 13; (d) faces of an individual with different expressions and hair occlusions. #feature matches: 41, #false matches: 8. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

RoPS feature \hat{f}_i is calculated as

$$\hat{f}_i = \mathbf{V}_{sf}^T \mathbf{f}_i. \quad (7)$$

3.2. 3D feature matching

Assume that $\mathcal{F}^i = \{f_n^i\}$ and $\mathcal{F}^j = \{f_m^j\}$ are the sets of RoPS features extracted from 3D faces \mathbf{P}^i and \mathbf{P}^j , respectively. The Nearest Neighbor Distance Ratio (NNDR) approach 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:

$$f_{m'}^j = \arg \min_{f_m^j \in \mathcal{F}^j} \|f_n^i - f_m^j\|_2, \quad (8)$$

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

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:

$$r_{dis} = \frac{\|f_n^i - f_{m'}^j\|_2}{\|f_n^i - f_{m''}^j\|_2}. \quad (10)$$

If the ratio r_{dis} is less than a threshold τ_f , $(f_n^i, f_{m'}^j)$ 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 $(f_n^i, f_{m'}^j)$ is finally considered as a feature match. The threshold τ_f determines both the number and accuracy of feature matches. A small threshold produces a limited number of feature matches and is not sufficient for 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. Fig. 5 presents an illustration of 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 4 mm, the feature

match is considered correct, and is denoted by a green line in Fig. 5. Otherwise, the feature match is considered false, and is denoted by a red line 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 results can be used to identify the same individual with different expressions, and to distinguish different individuals with the same expression.

We match all of the features in \mathcal{F}^i against the features in \mathcal{F}^j , resulting in a set of matched keypoints $\mathcal{C}^{ij} = \{c_1^{ij}, c_2^{ij}, \dots, c_{N_c}^{ij}\}$, where $c_n^{ij} = \{q_n^i, q_n^j\}$ 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 use the Procrustes algorithm [13] to perform registration between the two 3D faces \mathbf{P}^i and \mathbf{P}^j . The number of closest point pairs with distances less than 2 mm 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 [26] and the RANSAC algorithm [6,7]. In order to improve the computational efficiency, we only consider the aforementioned three metrics.

4. 3D face recognition

Face recognition includes two different tasks: face identification and face verification [4]. 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 \mathbf{s}_k can be obtained ($k = 1, 2, 3$). The m th element s_{km} in \mathbf{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 \mathbf{s}_k have to be normalized to the range of [0,1] using the min-max rule. That is,

$$\hat{\mathbf{s}}_k = \frac{\mathbf{s}_k - \min(\mathbf{s}_k)}{\max(\mathbf{s}_k - \min(\mathbf{s}_k)) - \min(\mathbf{s}_k - \min(\mathbf{s}_k))}. \quad (11)$$

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

$$\hat{\mathbf{s}}_2 = 1 - \hat{\mathbf{s}}_2. \quad (12)$$

Once these similarities are calculated, a fused similarity is calculated as:

$$\mathbf{s} = \sum_{k=1}^3 \omega_k \hat{\mathbf{s}}_k. \quad (13)$$

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:

$$\hat{\mathbf{s}} = \frac{\mathbf{s} - \min(\mathbf{s})}{\max(\mathbf{s} - \min(\mathbf{s})) - \min(\mathbf{s} - \min(\mathbf{s}))}. \quad (14)$$

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., mouth 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 two publically available datasets (i.e., the FRGC dataset [29] and the Bosphorus dataset [32]) to test our proposed E13D algorithm.

The FRGC 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 [7]. 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.

The Bosphorus dataset includes 4666 3D facial scans acquired from 105 individuals (61 men and 44 women) aged from 25 to 35

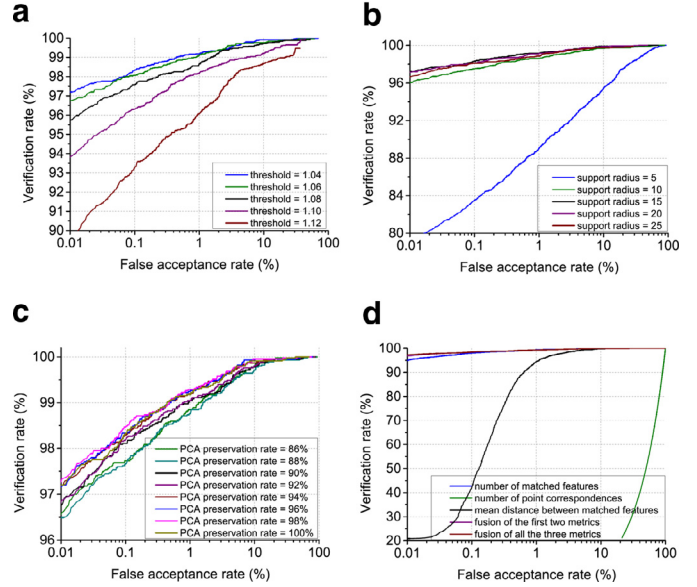


Fig. 6. Face verification results on the FRGC v1 dataset under different parameter settings. (a) Results with different thresholds of keypoint detection. (b) Results with different support radii for feature description. (c) Results with different preservation rates for feature compression. (d) Results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

[32]. There are more than 31 scans for each individual, these scans were acquired under different expressions, poses, and occlusions.

5.1.2. Evaluation criteria

For face identification, we adopt the frequently used Cumulative Match Characteristic (CMC) and Rank-1 Identification Rate (R1IR) 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 R1IR 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 [4].

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 15 mm, 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

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 |

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 being set to 5 mm, 10 mm, 15 mm, 20 mm and 25 mm. 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 5 mm to 10 mm. 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 10 mm to 15 mm, 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 15 mm. 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 15 mm.

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 being 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 being 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 has been preserved, and the un-

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 |

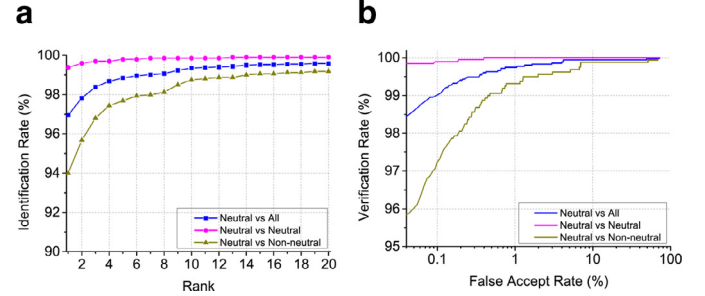


Fig. 7. Face recognition results on the FRGC v2 dataset. (a) CMC results for face identification. (b) ROC results for face verification.

desired 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.

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.

5.5. Comparative results on the FRGC v2 dataset

The FRGC v2 dataset was used to test the face verification and identification performance of our proposed algorithm. The fusion of metrics n_{fc} and n_{pc} was used in these experiments.

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

Table 5

Comparison of verification rates at 0.1% FAR achieved on the FRGC v2 dataset. The results shown in *italic* are achieved by algorithms following the keypoint-descriptor-matching framework. ‘NvsA’, ‘NvsN’, and ‘NvsNN’ stand for ‘Neutral vs All’, ‘Neutral vs Neutral’, and ‘Neutral vs Non-neutral’, respectively.

| Method | NvsA (%) | NvsN (%) | NvsNN (%) |
|-----------------------------------|--------------|-------------|-------------|
| <i>El3D</i> | 99.01 | 99.9 | 97.18 |
| <i>Elaiwat et al. [12], 2015</i> | NA | 99.6 | 93.1 |
| <i>Huang et al. [18], 2012</i> | 98.4 | 99.6 | 97.2 |
| <i>Lei et al. [21], 2015</i> | 98.3 | 99.9 | 96 |
| <i>Berretti et al. [5], 2010</i> | 95.5 | 97.7 | 91.4 |
| <i>Wang et al. [35], 2010</i> | 98.61 | 99.2 | 97.7 |
| <i>Al-Osaimi et al. [1], 2009</i> | 98.14 | 98.35 | 97.8 |
| <i>Mian et al. [26], 2008</i> | 97.4 | 99.9 | 92.7 |
| <i>Mian et al. [25], 2007</i> | 98.5 | 99.4 | 97.0 |

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 of the existing algorithms, achieving a high verification rate of 99.01%.

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 have a high repeatability. 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.

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 identification 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 identification 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 identification rate for all 3D faces is 97.0%.

To compare our identification results with the state-of-the-art performance achieved on the FRGC v2 dataset, we present the rank-1 identification rates of existing algorithms in Table 6. It can be observed that the best identification result is achieved by Queirolo et al. [30] and Wang et al. [35], with a rank-1 identification rate of 98.4%. The rank-1 identification rate achieved by our algorithm is 97.0%. Note that, the method presented in [30] uses a complicated surface registration process for face recognition, which is very time-consuming.

Table 6

Comparison of rank-1 identification rates achieved on the FRGC v2 dataset. The results shown in *italic* are achieved by algorithms following the keypoint-descriptor-matching framework.

| Method | Rank-1 identification rate (%) |
|-----------------------------------|--------------------------------|
| <i>El3D</i> | 97.0 |
| <i>Elaiwat et al. [12], 2015</i> | 97.1 |
| <i>Li et al. [23], 2014</i> | 96.3 |
| <i>Smeets et al. [34], 2013</i> | 89.6 |
| <i>Huang et al. [18], 2012</i> | 97.6 |
| <i>Mian et al. [26], 2008</i> | 93.5 |
| <i>Berretti et al. [5], 2010</i> | 94 |
| <i>Queirolo et al. [30], 2010</i> | 98.4 |
| <i>Alyuz et al. [3], 2010</i> | 97.5 |
| <i>Wang et al. [35], 2010</i> | 98.4 |
| <i>Al-Osaimi et al. [1], 2009</i> | 96.5 |

5.5.3. Discussions

Several algorithms following the same framework (i.e., the keypoint-descriptor-matching framework) as our algorithm can also be found in the literature [12,18,23,26,34], their results are shown in Table 5 and 6 in *italic*. The significant advantage of this framework is its strong robustness to occlusion, clutter and pose variations. It can be observed that our algorithm achieves the best face verification performance among these algorithms, with a VR@0.1%FAR of more than 99%. Besides, our face identification performance is close to the best result reported by Huang et al. [18]. Particularly, Smeets et al. [34] used Mesh-SIFT features for facial representation and achieved a rank-1 identification rate of 89.6%. Elaiwat et al. used curvelet to detect keypoints and extract facial descriptors, and achieved a rank-1 identification rate of 97.1%. Li et al. [23] proposed a Mesh-Daisy detector and three descriptors including the Histogram of Gradient (HOG), the Histogram of Shape index (HOS), and the Histogram of Gradient of Shape index (HOGS). A rank-1 identification rate of 96.3% was achieved by Li et al. [23]. Besides, Huang et al. used eLBP for 3D facial representation and achieved the highest rank-1 identification rate of 97.6%. When comparing all of the results shown in Table 5 and 6, it can also be interpreted that our algorithm is better designed for face verification rather than face identification. Its face verification performance is superior to existing algorithms.

5.6. Comparative results on the Bosphorus dataset

The Bosphorus dataset was further used to test the face identification performance of our proposed algorithm. First, keypoints were detected using the technique described in Section 2.2. For each keypoint, its neighboring points (with a radius of 15 mm) were transformed to its local reference frame for RoPS feature generation [15,31]. These RoPS descriptors were then compressed and matched using the technique presented in Section 3. Finally, face recognition was performed using the similarity metric η_{fc} (Section 4). Our algorithm was tested on the ‘Occlusion’ subset of the Bosphorus dataset. The ‘Occlusion’ subset consist of 381 3D facial scans with occlusions of eyes by hand (E-Hand), occlusions of mouth by hand (M-Hand), occlusions of eyes by glasses (E-Glasses) and occlusions of facial regions by hair (F-Hair).

To perform rigorous comparison on the dataset, we follow the same experimental setup as [11,21]. That is, the gallery is generated by selecting a neutral 3D face from each individual, the probe set consists of 381 3D faces with external occlusions (i.e., E-hand, M-Hand, E-Glasses, F-Hair). The CMC results are shown in Fig. 8. It can be observed that our algorithm achieved a high identification rate under different occlusions. For the faces with occlusions of E-Glasses, the rank-1 identification rate is as high as 99.04%, the rank-2 identification rate is increased to 100%. These

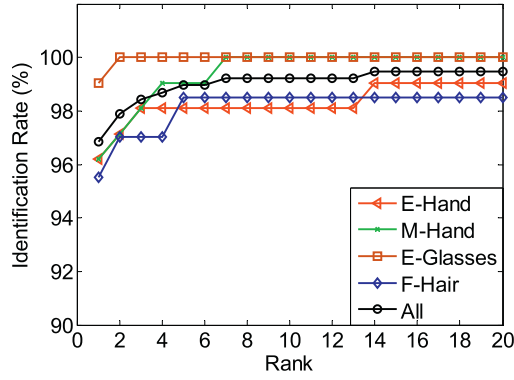


Fig. 8. The CMC results achieved on the ‘Occlusion’ subset of the Bosphorus dataset.

Table 7

Comparison of the rank-1 identification rates achieved on the Bosphorus dataset.

| | E-hand (%) | M-hand (%) | E-glasses (%) | F-hair (%) | Overall (%) |
|---------------------------|---------------|---------------|------------------|---------------|----------------|
| EI3D | 96.19 | 96.19 | 99.04 | 95.52 | 96.85 |
| Lei et al. [21], 2015 | 90.5 | 94.3 | 96.2 | 88.1 | 92.7 |
| Berretti et al. [7], 2014 | NA | NA | NA | NA | 95.8 |
| Li et al. [23], 2014 | 100 | 100 | 100 | 95.5 | 99.2 |
| Berretti et al. [6], 2013 | NA | NA | NA | NA | 93.2 |
| Drira et al. [11], 2013 | 97.1 | 74.7 | 94.2 | 81 | 87 |
| Colombo et al. [10], 2011 | 91.1 | 74.7 | 94.2 | 90.4 | 87.6 |
| Alyuz et al. [2], 2008 | 93.6 | 93.6 | 97.8 | 89.6 | 93.6 |
| SHOT based | 91.4 | 91.4 | 90.4 | 92.5 | 91.3 |

results indicates that our algorithm is very robust to occlusions caused by glasses. For faces with occlusions of E-Hand, M-Hand, and F-Hair, their rank-1 identification rates are 96.19%, 96.19% and 95.52%, respectively. The overall rank-1 identification rate for all 3D faces is 96.85%.

The rank-1 identification rates of our algorithm were further compared with the state-of-the-art, as shown in Table 7. Our algorithm achieves the second best performance, with an overall rank-1 identification rate of 96.85%. It outperforms the methods [6,10,11,21] and [2] by a large margin. The best performance was achieved by Li et al. [23], with an overall rank-1 identification rate of 99.2%. Note that, Li et al. combined three histogram based features to generate the Histogram of Multiple surface differential Quantities (HOMQ) for 3D face representation, and then used the Sparse Representation based Classifier (SRC) for face recognition. Both of the complicated feature fusion and the fine-grained feature matching stages have contributed to the high performance of [23]. Considering each individual feature and the coarse-grained matcher (which is similar to our algorithm), the rank-1 identification rates achieved by HOG, HOS and HOGS are 92.65%, 97.38% and 93.96%, respectively. These results are inferior or slightly better than our results. This indicates that the performance of our algorithm can still be significantly improved by using feature fusion and more advanced classifiers.

To further test the effectiveness of our RoPS feature descriptor, we replaced the RoPS descriptor with the popular Signature of Histograms of Orientations (SHOT) descriptor [31]. The other parts of our algorithm remain the same as those described in Section 5.6. The results achieved by the SHOT based algorithm is given in Table 7. It can be observed that the overall identification rate achieved by the RoPS based algorithm is 91.3%, which is inferior to the results achieved by the RoPS based algorithm. It can be interpreted that the RoPS feature is better than the SHOT feature. This clearly demonstrates the effectiveness of our RoPS descriptor.

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 and Bosphorus datasets show that the proposed algorithm achieves high face identification and verification rates. Moreover, our algorithm is robust to expression variations and occlusions. Please also note that, the proposed 3D face recognition algorithm highly relies on accurate nosetip detection, therefore, developing more advanced nosetip-free 3D face recognition algorithm would be our future work.

Acknowledgments

This research is supported by grants from National Natural Science Foundation of China (nos. 61471371, 61403265, and 61202336) and Australian Research Council (DE120102960, DP150100294 and DP110103336). This work is also partly supported by the Science and Technology Plan of Sichuan Province (no. 2015SZ0226).

References

- [1] F. Al-Osaimi, M. Bennamoun, A. Mian, An expression deformation approach to non-rigid 3D face recognition, *Int. J. Comput. Vis.* 81 (3) (2009) 302–316.
- [2] N. Alyuz, B. Gokberk, L. Akarun, A 3d face recognition system for expression and occlusion invariance, in: *Proceedings of the Second IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2008, pp. 1–7.
- [3] N. Alyuz, B. Gokberk, L. Akarun, Regional registration for expression resistant 3-D face recognition, *IEEE Trans. Inf. Forensics Secur.* 5 (3) (2010) 425–440.
- [4] M. Bennamoun, Y. Guo, F. Sohel, Feature selection for 2D and 3D face recognition, *Encyclopedia of Electrical and Electronics Engineering*, John Wiley and Sons, Inc., 2015, pp. 1–54.
- [5] S. Berretti, A. Del Bimbo, P. Pala, 3D face recognition using isogeodesic stripes, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (12) (2010) 2162–2177.
- [6] S. Berretti, N. Werghi, A. del Bimbo, P. Pala, Matching 3D face scans using interestpoints and local histogram descriptors, *Comput. Graph.* (2013).
- [7] S. Berretti, N. Werghi, A. del Bimbo, P. Pala, Selecting stable keypoints and local descriptors for person identification using 3D face scans, *Vis. Comput.* 30 (11) (2014) 1275–1292.
- [8] A.M. Bronstein, M.M. Bronstein, R. Kimmel, Expression-invariant representations of faces, *IEEE Trans. Image Process.* 16 (1) (2007) 188–197.
- [9] C.S. Chua, F. Han, Y.K. Ho, 3D human face recognition using point signature, in: *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 233–238.
- [10] A. Colombo, C. Cusano, R. Schettini, Three-dimensional occlusion detection and restoration of partially occluded faces, *J. Math. Imaging Vis.* 40 (1) (2011) 105–119.
- [11] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, R. Slama, 3D face recognition under expressions, occlusions and pose variations, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (9) (2013) 2270–2283.
- [12] S. Elaiwat, M. Bennamoun, F. Boussaid, A. El-Sallam, A curvelet-based approach for textured 3D face recognition, *Pattern Recognit.* 48 (4) (2015) 1235–1246.
- [13] C. Goodall, Procrustes methods in the statistical analysis of shape, *J. R. Stat. Soc. Ser. B Methodol.* (1991) 285–339.
- [14] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, 3D object recognition in cluttered scenes with local surface features: a survey, *IEEE Trans. Pattern Anal. Mach. Intell.* 36 (11) (2014a) 2270–2287.
- [15] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, J. Wan, Rotational projection statistics for 3D local surface description and object recognition, *Int. J. Comput. Vis.* 105 (1) (2013) 63–86.
- [16] Y. Guo, F. Sohel, M. Bennamoun, J. Wan, M. Lu, An accurate and robust range image registration algorithm for 3D object modeling, *IEEE Trans. Multimed.* 16 (5) (2014b) 1377–1390.
- [17] S. Gupta, M.K. Markey, A.C. Bovik, Anthropometric 3D face recognition, *Int. J. Comput. Vis.* 90 (3) (2010) 331–349.
- [18] D. Huang, M. Ardabilian, Y. Wang, L. Chen, 3-D face recognition using eLBP-based facial description and local feature hybrid matching, *IEEE Trans. Inf. Forensics Secur.* 7 (5) (2012) 1551–1565.

- [19] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, Local binary patterns and its application to facial image analysis: a survey, *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 41 (6) (2011) 765–781.
- [20] Y. Lei, M. Bennamoun, M. Hayat, Y. Guo, An efficient 3D face recognition approach using local geometrical signatures, *Pattern Recognit.* 47 (2) (2014) 509–524.
- [21] Y. Lei, Y. Guo, M. Hayat, M. Bennamoun, X. Zhou, A two-phase weighted collaborative representation for 3d partial face recognition with single sample, *Pattern Recognit.* 52 (2016) 218–237.
- [22] H. Li, D. Huang, J.-M. Morvan, L. Chen, Y. Wang, Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns, *Neurocomputing* 133 (2014) 179–193.
- [23] H. Li, D. Huang, J.-M. Morvan, Y. Wang, L. Chen, Towards 3D face recognition in the real: a registration-free approach using fine-grained matching of 3D key-point descriptors, *Int. J. Comput. Vis.* 113 (2) (2015) 128–142.
- [24] P. Liu, Y. Wang, D. Huang, Z. Zhang, L. Chen, Learning the spherical harmonic features for 3-D face recognition, *IEEE Trans. Image Process.* 22 (3) (2013) 914–925.
- [25] A. Mian, M. Bennamoun, R. Owens, An efficient multimodal 2D-3D hybrid approach to automatic face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (11) (2007) 1927–1943.
- [26] A. Mian, M. Bennamoun, R. Owens, Keypoint detection and local feature matching for textured 3D face recognition, *Int. J. Comput. Vis.* 79 (1) (2008) 1–12.
- [27] H. Mohammadzade, D. Hatzinakos, Iterative closest normal point for 3D face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (2) (2013) 381–397.
- [28] G. Passalis, P. Perakis, T. Theoharis, I.A. Kakadiaris, Using facial symmetry to handle pose variations in real-world 3D face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (10) (2011) 1938–1951.
- [29] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1, 2005, pp. 947–954.
- [30] C.C. Queirolo, L. Silva, O. Bellon, M.P. Segundo, 3D face recognition using simulated annealing and the surface interpenetration measure, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (2) (2010) 206–219.
- [31] S. Salti, F. Tombari, L.D. Stefano, SHOT: unique signatures of histograms for surface and texture description, *Comput. Vis. Image Underst.* 125 (8) (2014) 251–264.
- [32] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, Bosphorus database for 3d face analysis, in: *Biometrics and Identity Management*, Springer, 2008, pp. 47–56.
- [33] D. Smeets, P. Claes, J. Hermans, D. Vandermeulen, P. Suetens, A comparative study of 3-D face recognition under expression variations, *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 42 (5) (2012) 710–727.
- [34] D. Smeets, J. Keustermans, D. Vandermeulen, P. Suetens, meshSIFT: local surface features for 3D face recognition under expression variations and partial data, *Comput. Vis. Image Underst.* 117 (2) (2013) 158–169.
- [35] Y. Wang, J. Liu, X. Tang, Robust 3D face recognition by local shape difference boosting, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (10) (2010) 1858–1870.
- [36] H.-S. Wong, K. Cheung, H. Ip, 3D head model classification by evolutionary optimization of the Extended Gaussian Image representation, *Pattern Recognit.* 37 (12) (2004) 2307–2322.