Pattern Recognition 52 (2016) 218–237

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A Two-Phase Weighted Collaborative Representation for 3D partial face recognition with single sample

Yinjie Lei [a], Yulan Guo [b], Munawar Hayat [c,d], Mohammed Bennamoun [c], Xinzhi Zhou [a,*]

[a] College of Electronics and Information Engineering, Sichuan University, Chengdu, Sichuan, China
[b] College of Electronic Science and Engineering, National University of Defense Technology, Changsha, Hunan, China
[c] School of Computer Science and Software Engineering, University of Western Australia, Crawley, WA, Australia
[d] IBM Research Australia, Carlton, VIC, Australia

ARTICLE INFO

Article history:
Received 25 March 2015
Received in revised form
30 June 2015
Accepted 30 September 2015
Available online 23 October 2015

Keywords:
3D face recognition
3D representation
Sparse representation
Partial facial data
Single sample problem

ABSTRACT

3D face recognition with the availability of only partial data (missing parts, occlusions and data corruptions) and single training sample is a highly challenging task. This paper presents an efficient 3D face recognition approach to address this challenge. We represent a facial scan with a set of local Keypoint-based Multiple Triangle Statistics (KMTS), which is robust to partial facial data, large facial expressions and pose variations. To address the single sample problem, we then propose a Two-Phase Weighted Collaborative Representation Classification (TPWCRC) framework. A class-based probability estimation is first calculated based on the extracted local descriptors as a prior knowledge. The resulting class-based probability estimation is then incorporated into the proposed classification framework as a locality constraint to further enhance its discriminating power. Experimental results on six challenging 3D facial datasets show that the proposed KMTS–TPWCRC framework achieves promising results for human face recognition with missing parts, occlusions, data corruptions, expressions and pose variations.

1. Introduction

Face recognition (FR) is an active research area in the computer vision community due to its non-intrusive and friendly acquisition nature when compared to other biometrics. The task of FR is to identify or verify a human face from its records (2D and 3D modalities). FR has a number of real-world applications including access control and video surveillance. Among various facial modalities, considerable progress has been made with 2D FR. However, its performance is still challenged by pose changes and illumination variations. With the rapid development of 3D acquisition technologies [1–7], 3D FR has drawn growing attention due to its potential capability to overcome the inherent limitations of its 2D counterpart. Most of the existing 3D FR approaches are proposed for facial scans acquired in a highly controlled environments. Such facial scans are mostly frontal and good in quality. Promising recognition performance has been reported on those high quality 3D facial data by existing 3D FR approaches. However, for many real-world applications, facial scans can only be acquired in uncontrolled environments. Such facial scans usually contain a partial face in the presence of missing parts, occlusions and data corruptions (as shown in Fig. 1). FR from such partial facial scans is still an open issue, and further investigation is required to enable fully automatic 3D Partial Face Recognition (PFR). Although a few 3D PFR approaches have been proposed, the recent release of datasets containing partial facial scans (e.g., the Bosphorus [8], GavabDB [9] and UMB-DB [10] datasets) provides a large benchmark for 3D PFR and has further boosted the research on this topic. Furthermore, most of the existing 2D/3D FR approaches require a sufficiently large set of training samples per individual to cover possible facial variations (e.g., partial data and expressions) for accurate FR. However, in many real-world FR applications, only a single sample per individual can be provided for training, resulting in the single sample based FR problem. Partial facial data and the limitation in the number of training samples per individual are therefore two important challenges for real-world 3D PFR applications.

The existing 3D FR approaches can be coarsely classified into two categories: global descriptor based and local descriptor based approaches. The global descriptor based approaches extract features from the entire face to encode the geometric characteristics of a 3D face [5,7]. However, these approaches only work when the complete 3D face is available. These approaches therefore rely on the availability of complete facial scan and are sensitive to missing parts,

* Corresponding author.
E-mail addresses: yinjie@scu.edu.cn (Y. Lei), yulan.guo@nudt.edu.cn (Y. Guo), hayatm@au1.ibm.com (M. Hayat), mohammed.bennamoun@uwa.edu.au (M. Bennamoun), xz.zhou@scu.edu.cn (X. Zhou).

http://dx.doi.org/10.1016/j.patcog.2015.09.035
0031-3203/

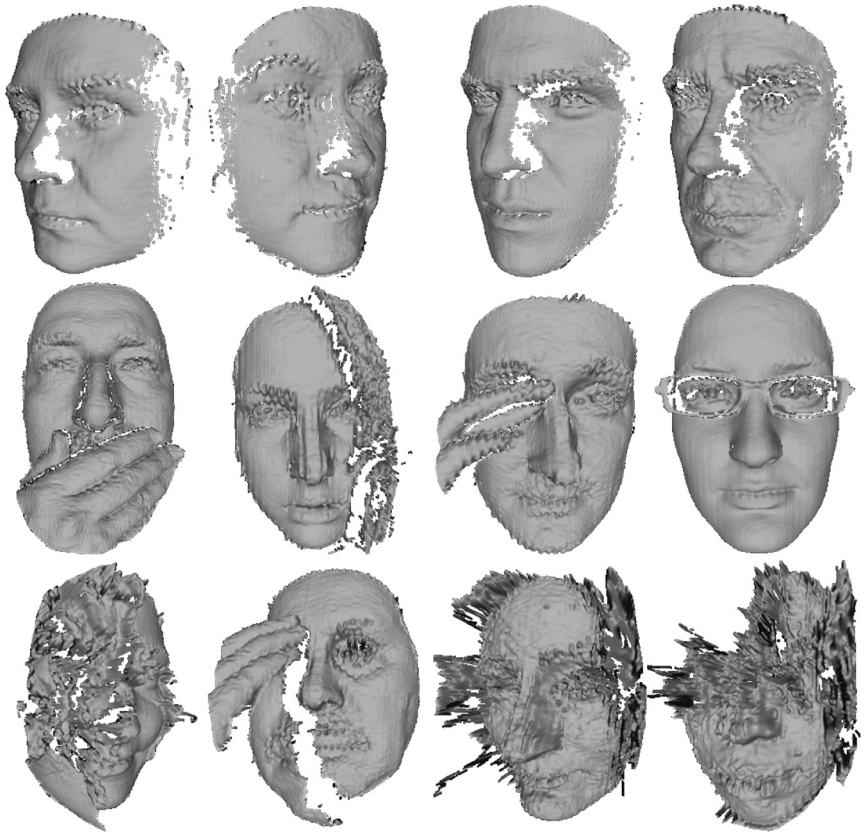

Fig. 1. Major challenges for 3D PFR: missing parts (first row), occlusions (second row) and data corruptions (third row).

occlusions and data corruptions. In contrast, local descriptor based approaches usually extract features from the neighborhoods of the detected keypoints, resulting in a set of coordinate-independent local descriptors. Compared to global descriptor based approaches, the number of extracted local descriptors is related to the content of input face (holistic or partial). Therefore, such flexible facial representation is more suitable for 3D PFR. Furthermore, it is commonly believed that only a few facial regions are significantly affected by distortions caused by missing parts, occlusions and data corruptions, while most of the other regions remain invariant [11,12]. This indicates that local facial descriptors are more suitable and powerful when dealing with partial facial data.

Recently, Sparse Representation (SR) has become a powerful method for various pattern recognition tasks. Experimental results on FR show that the Sparse Representation-based Classification (SRC) algorithms outperform a number of conventional FR algorithms. However, solving l_1-norm based sparsity usually requires computationally demanding optimization procedures. It is argued that the l_1-norm sparsity is not essential for the improvement of FR performance [13]. Several more efficient algorithms have therefore been proposed, including the Collaborative Representation-based Classification (CRC), which can achieve similar recognition performance compared to SRC [13,14]. However, both SRC and CRC require multiple training samples for each individual to cope with the test input with complex variations. In the case of single sample problem, the representation becomes unreliable due to the lack of training samples. Inspired by this fact, more prior classification knowledge can be incorporated into the data representation to perform single sample based FR.

In this paper, we propose a Two-Phase Weighted Collaborative Representation Classification (TPWCRC) for single sample based 3D PFR. First, an efficient Keypoint-based Multiple Triangle Statistics (KMTS) descriptor is presented for facial scan representation (Section 3). The proposed KMTS descriptor exhibits both high discriminative power and robustness under various deformations (i.e., partial data, expressions and pose variations). Second, a class-based probability estimation algorithm is presented to provide a strong prior classification knowledge for any input test during the first phase of classification. The probability estimation can then be used to compensate for the unavailability of multiple training samples per individual. The resulting probability estimation is therefore integrated into the TPWCRC framework as a locality constraint to restrain the globality of the data representation of the input test. The final classification result then corresponds to the individual which gives the smallest residuals (Section 4). The proposed KMTS–TPWCRC 3D PFR approach is alignment-free and no manual intervention is required. Its efficiency and robustness has been extensively demonstrated by a set of experiments on six popular 3D face datasets (Section 5).

Furthermore, the proposed approach only involves simple computations, including nosetip detection, keypoint detection, local geometrical descriptor and a TPWCRC framework which can be derived analytically. For the considerations of developing an efficient 3D PFR system, simple geometric features are extracted without any complex mathematical operations. Unlike many existing 3D PFR approaches which rely on complicated and time-consuming feature extraction methods (e.g., [15,16]) or occlusion data restoration method (e.g., [15,17]), the proposed approach is very efficient (see Section 5.5). Its efficiency can further be improved by performing these simple computations in parallel. Although some existing methods achieve better FR performance, the proposed approach can achieve a comparable performance at much lower computational cost. This suggests that our proposed approach is more suitable for practical applications.

The rest of this paper is organized as follows. Section 2 provides a brief literature review of closely related 3D FR approaches.

Section 3 describes the proposed KMTS facial descriptors. Section 4 introduces the proposed TPWCRC framework for single sample based 3D PFR. Section 5 presents the experimental results and comparative analysis on five popular 3D facial datasets. Section 6 concludes the paper.

2. Related work and our contributions

FR has been well studied and several survey papers can be found in [18,19]. In this section, we first briefly summarize the existing literature which is most relevant to our approach. Specifically, we will restrain our review to local descriptor based 3D FR approaches (Section 2.1), 3D PFR approaches (Section 2.2), and SRC based facial analysis approaches (Section 2.3). We will then give an overview of the proposed approach in Section 2.4.

2.1. Local descriptor based 3D face recognition

Unlike holistic representation, local descriptor based 3D FR approaches detect and match the prominent points/regions of a face instead of full faces, in order to handle expressions and partial data and thus improve the recognition performance. However, most of the existing local descriptor based 3D FR approaches rely on the one-to-one matching between all possible pairs of the probe and gallery faces. Although some of them can perform face verification in nearly real-time, face identification is still a major challenge for those approaches, especially when the gallery size is large. Therefore, those approaches are not appropriate for practical applications.

Mian et al. [20] proposed a salient point detection algorithm for 3D facial scans, which can obtain a high repeatability. They fused 3D salient point features with 2D Scale Invariant Feature Transform (SIFT) features to develop a robust 3D+2D multi-modal FR system. Faltemier et al. [21] extracted a set of small facial subregions in order to improve their discriminative power for 3D FR. The resulting subregions were matched independently, and their corresponding matching scores were fused. The highest recognition performance was achieved by selecting and combining 28 prominent subregions. Similarly, Chang et al. [22] segmented each 3D facial scan into multiple subregions. A set of subregions located around the nose were then individually matched to generate the final matching score. Gupta et al. [23] automatically detected 10 anthropometric fiducial points on a 3D facial surface. Then, a stochastic method was used to calculate both the 3D Euclidean and geodesic distances between all pair-wise combinations of the fiducial points. Their results showed that the geodesic distance was more reliable in terms of recognition accuracy compared to the Euclidean distance. Lei et al. [11] divided a 3D facial scan into rigid, semi-rigid and non-rigid subregions. A set of local geometrical descriptors extracted from the rigid and semi-rigid subregions were fused. Then, the resulting descriptors were fed into a Support Vector Machine (SVM) classifier to perform 3D FR. Li et al. [24] designed various types of low-level geometric features (such as curvatures) for 3D facial surface. Then, a feature pooling and ranking algorithm was deployed to select the features which are robust to expressions. Wang et al. [5] first accurately aligned two 3D facial scans. Then, a signed shape-difference map was generated as an intermediate representation for facial shape comparison and matching. Ballihi et al. [25] used the theory of Riemannian geometry to define a set of geodesic paths between nasal curves of different facial surfaces. The length of the geodesic path was used as a similarity measure. Then, the AdaBoost algorithm was used to select the most discriminative nasal curves. Lei et al. [12] proposed a novel feature descriptor called Angular Radial Signatures (ARSs) from the semi-rigid region of a 3D face. The resulting descriptor is fed into an SVM classifier to perform 3D FR. Berretti et al. [26] represented a 3D face with a set of isogeodesic stripes. A pair of corresponding stripes was matched using the 3D Weighted Walkthroughs (3DWWs) algorithm. Queirolo et al. [27] used the local Surface Inter-penetration Measure (SIM) to match 3D faces. The authentication score was obtained by combining the SIM values corresponding to four different facial regions (i.e., circular and elliptical areas around the nose, facial forehead, and the entire face).

2.2. Partial 3D face recognition

3D FR from high quality facial scans (acquired in highly controlled environment) has been extensively explored in recent years. However, 3D FR from low quality scans is still in its infancy and only a few 3D FR approaches consider partial facial data. Although the related 3D PFR approaches can deal with the missing parts, occlusions and data corruptions, some approaches require a time-consuming data restoration process when dealing with data corruptions. The other approaches need manual interference to partition facial scan into occluded/unoccluded regions. Thus, they are not suitable for efficient real-world applications.

Alyuz et al. [28] used the Average Region Models (ARMs) to handle occlusions and missing parts. They manually segmented the facial surface into several meaningful subregions and the registration of faces was performed by separate dense alignments with the corresponding ARMs. However, one major limitation of the approach is the intensive manual efforts required for subregion segmentation. Berretti et al. [29] represented a 3D face with multiple meshDOG keypoints and local geometric histogram descriptors. Then, they selected the most effective features from the local descriptors to perform partial face matching. Similarly, Drira et al. [15] extracted radial curves which emanate from the nosetip to represent facial surfaces. Then, a Riemannian framework was developed to analyze these curves. Their approach can effectively handle expressions, pose variations, missing data and occlusions. Colombo et al. [30] proposed an occlusion detection scheme by analyzing the difference between an original face and its eigenface approximation. The regions detected as occlusions were removed and restored using the Gappy PCA algorithm. Then, face identification was performed following the Fisherfaces algorithm. Alyuz et al. [17] proposed a fully automatic 3D face recognition system which is robust to occlusions. They first employed an adaptively selected model based registration scheme to detect and remove occlusions. Then, a masking strategy was proposed in the classification stage which used the subspace analysis techniques for incomplete data. Colombo et al. [31] first roughly detected the facial occlusions using the difference between a face and the mean face, followed by a restoration process. The Gappy PCA was then used to discriminate face from non-face images when only partial facial data is available. Li et al. [32] proposed a preprocessing algorithm to obtain a canonical frontal view of a facial surface, which is independent to their initial poses (especially for profile faces). The resulting canonical face was approximated by sparse coding, and the query face was then assigned to the identity with the smallest reconstruction error. Passalis et al. [33] proposed an Annotated Face Model (AFM) to register and fit a facial scan. The facial symmetry was used to overcome the challenges raised by missing parts during fitting. Comparisons were performed among interpose scans using a wavelet-based biometric signature. Zhao et al. [34] proposed a statistical facial feature model for landmark localization in 3D faces under expressions and occlusions, which was then used for 3D PFR. In order to deal with occlusions, an occlusion classifier and a fitting algorithm were introduced. Chu et al. [35] proposed a semi-coupled dictionary learning algorithm by computing a jointly-optimized solution with incorporated

reconstruction cost, which can be applied for 3D PFR. Li et al. [16] proposed a 3D PFR approach using a curvature-based 3D keypoint detector and three sophisticated keypoint descriptors. Then, a multi-task sparse representation based fine-grained matching algorithm is applied to perform face recognition. Their proposed approach achieved the best recognition performance on the Bosphorus dataset.

2.3. SRC based face recognition

Several existing SRC based facial analysis approaches (using both 2D and 3D modalities) can be found in the literature. Typical SRC based approaches usually work in the scenario with a rich set of training images for each individual. In that case, an input face can easily be reconstructed from this rich training set during the test phase. However, many real-world FR applications can only offer single training face per individual. Consequently, the typical SRC based FR approaches become no longer reliable when dealing with the single sample based 3D FR problem.

Wright et al. [36] introduced the well-known SRC algorithm for 2D FR. It was demonstrated that the choice of feature space is no longer critical. The SRC with random projection-based features outperformed a number of conventional face recognition schemes. LBP-SRC [37] used SRC in conjunction with multi-resolution histograms based on the Local Binary Patterns (LBP). The approach is robust to geometric misalignment between the gallery and query faces. Wagner et al. [38] proposed an improved SRC algorithm to consider the registration errors. In order to cater for images acquired under uncontrolled illuminations, they presented an image illumination synthesis scheme from a set of training samples. Liao et al. [39] proposed an alignment-free approach to address the problem of 2D PFR. A face image was represented with a set of Gabor ternary pattern descriptors extracted from the detected SIFT keypoints. Then, each keypoint was fed into an individual multi-task SRC for classification. Similarly, Zhang et al. [40] proposed a 3D FR approach by multiple meshSIFT keypoint based descriptors and SRC. The approach does not require any pre-alignment between two facial scans and is robust to various challenging conditions. Deng et al. [41] proposed an Extended SRC to handle the problem of insufficient training samples. The Extended SRC algorithm uses an auxiliary intraclass variant dictionary to represent the possible variations between the training and test samples. The proposed approach can even generalize to the case of single training sample per class. Motivated by the fact that kernel trick can capture the nonlinear similarity of features, Gao et al. [42] proposed Kernel Sparse Representation (KSR), which is indeed the SRC in a high dimensional feature space mapped by a kernel function. Their experimental results showed that KSR outperformed typical SRC algorithms for both object and face recognition. Xu et al. [43] presented a two-phase test sample representation for FR. During the first phase, the test sample was represented with the linear combination of all the training samples via SRC. Then, the representation capability of each training sample was exploited to determine M nearest neighbors for the test sample, which was used to linearly represent the test sample during the second classification phase. Yang et al. [44] introduced a Gabor occlusion dictionary computing algorithm to obtain a more compact occlusion dictionary. Their proposed approach reduces the computational cost in coding the occluded face images and significantly improves the accuracy of original SRC algorithm.

2.4. Contributions of this work

In this paper, we propose an efficient and robust framework which can effectively deal with single sample-based 3D PFR using a novel KMTS facial representation and a TPWCRC classification scheme. The major contributions of this paper can be summarized as follows:

- We present a discriminative and robust local facial descriptor called KMTS (Section 3). KMTS can effectively handle expressions, pose variations and other rigid transformations while being computationally efficient. Furthermore, the size of each KMTS descriptor is determined by the specific content of the input face, making KMTS highly suitable for 3D PFR.
- We propose a TPWCRC framework to perform single sample based 3D PFR (Section 4). The classification accuracy of the proposed TPWCRC framework is improved by incorporating a strong prior classification knowledge.
- We integrate the proposed KMTS and TPWCRC into a complete framework for single sample based 3D PFR. The proposed framework can work in the scenarios with single training sample and is adaptive to both holistic or partial facial data. The proposed framework is alignment-free and robust to facial expressions and pose variations. The proposed KMTS–TPWCRC framework is computationally efficient compared with the existing 3D PFR solutions, making it suitable for practical applications.
- Experimental results demonstrate that our approach achieves superior or comparable performance compared to the existing approaches. Furthermore, the proposed approach can effectively deal with occlusions, partial and missing data and is suitable for the application scenarios with limited availability of training samples (Section 5).

3. KMTS based 3D face representation

3.1. Motivation of the proposal of KMTS

Most of the existing 3D FR approaches rely on holistic facial descriptors with a predefined dimensionality. However, in the case of 3D PFR, a facial scan may suffer from missing parts, occlusions and data corruptions and it is not always feasible to extract fixed-length holistic feature descriptors. Therefore, in this work, we propose a local descriptor for 3D PFR based on our previous work [11,20]. Our descriptor is specifically tailored for 3D PFR with the following considerations. First, the number of the facial descriptors should be determined by the available content of the input facial scan. That is, a holistic facial scan will produce descriptors with a larger size than a partial one. Second, the local keypoints should be extracted with a high repeatability under data noise, pose variations and facial expressions. Third, the features extracted from these keypoints should be sufficiently distinctive in order to facilitate accurate FR. Last, the calculation of feature descriptors should be computationally efficient without any complicated mathematical transformations. This would make it suitable for rapid real-world 3D FR applications.

Our proposed KMTS descriptor consists of two parts. First, a set of robust keypoints is detected on a facial surface. The number of keypoints is flexible and is determined by the content of the input facial scan. Second, the multi-triangle statistics are extracted from the local neighborhood of each keypoint. The feature is directly computed from the spatial relationships of the 3D facial vertices and uses only some basic mathematical operations (e.g., distance and angle calculations). In the following subsections, the proposed KMTS descriptor is described in detail.

3.2. Keypoint detection

Given a facial range image $\mathbf{R}$ and its pointcloud $\mathbf{F}=[x_i, y_i, z_i]^T$, where $i=1,\ldots,n$, and n is the number of points. The 3D facial keypoints are detected using our previously proposed approach

[20]. The approach is used due to its superior performance in terms of repeatability, distinctiveness and robustness [45,46]. The keypoint detection procedure is briefly described below for completeness. These 3D keypoints can be repetitively detected on 3D facial scans under different variations. The high repeatability of the applied keypoint detection algorithm had been evaluated in [20] and a comprehensive survey paper [45]. For more details on the evaluation of the repeatability of the applied 3D keypoint detection algorithm, the reader should refer to [20,45]. Moreover, the introduced 3D facial keypoint detection algorithm is based on simple calculation and no complicated mathematical operations are involved.

First, **F** is sampled at uniform intervals of 4 mm in each direction of (x, y), and each sample point p is considered as a keypoint candidate for further computation. A local patch $\mathbf{L}=[x_j, y_j, z_j]^T$ centered at p is cropped by a sphere with radius r, where $j=1,\ldots,n_l$, and n_l is its number of patch points. The cropped patch will then be used to calculate the salience of the local neighborhood around p. The mean vector $\mathbf{m}$ and its corresponding covariance matrix $\mathbf{C}$ of patch $\mathbf{L}$ can be computed as

$$\mathbf{m}=\frac{1}{n_l}\sum_{j=1}^{n_l}\mathbf{L}_j, \quad \mathbf{C}=\frac{1}{n_l}\sum_{j=1}^{n_l}\mathbf{L}_j\mathbf{L}_j^T-\mathbf{m}\mathbf{m}^T \tag{1}$$

where $\mathbf{L}_j$ is the jth column of $\mathbf{L}$. Principal Component Analysis (PCA) is then performed on the covariance matrix $\mathbf{C}$ to obtain the diagonal matrix of the eigenvalues $\mathbf{D}$ and the corresponding eigenvector matrix $\mathbf{V}$:

$$\mathbf{CV}=\mathbf{DV} \tag{2}$$

The matrix $\mathbf{L}$ can be aligned with its principal axes using the Hotelling transform: $\mathbf{L}'_j=\mathbf{V}(\mathbf{L}_j-\mathbf{m})$. Let $\mathbf{L}_x$ and $\mathbf{L}_y$ be the x and y components of the patch $\mathbf{L}$, i.e, $\mathbf{L}'_x=x_j$ and $\mathbf{L}'_y=y_j$ (where $j=1,\ldots,n_l$).

$$\zeta=(\max(\mathbf{L}'_x)-\min(\mathbf{L}'_x))-(\max(\mathbf{L}'_y)-\min(\mathbf{L}'_y)) \tag{3}$$

In Eq. (3), ζ is the difference between the first two principal axes of the local patch. The value of ζ is zero if $\mathbf{L}'$ is planar or spherical. However, if the variation in the depth of $\mathbf{L}'$ is unsymmetrical, then ζ will have a non-zero value which is proportional to the variation. This depth variation provides an indication of the descriptiveness information contained in $\mathbf{L}$.

If the value of ζ is larger than a threshold (i.e., $\zeta \geq t$), p is considered as a keypoint, otherwise it is rejected. The threshold t determines the total number of detected keypoints. Increasing the value of t will decrease the number of keypoints. In this work, the values of r and t are empirically chosen as $r=20$ mm and $t=2$ mm. Fig. 2 shows an example of the detected keypoint and the local patch around a keypoint.

3.3. Multiple triangle statistics functions

Once a keypoint p is detected and its corresponding local patch $\mathbf{L}$ is extracted, we then represent each $\mathbf{L}$ with our proposed KMTS. First, the local surface $\mathbf{L}$ (centered at keypoint p) is represented by multiple spatial triangles, as shown in Fig. 3. For each triangle, one vertex is the keypoint p, and the other two vertices are randomly picked from $\mathbf{L}$. Four types of geometrical features are then calculated as follows:

- A: the angle between the two lines determined by the two random vertices and the keypoint.
- C: the radius of the circle circumscribed to the triangle determined by the two random vertices and the keypoint.
- D: the distance of the line between the two random vertices.
- N: the angle between the z-axis and the line determined by the two random vertices.

An illustration of the four geometrical features used to compute the proposed KMTS descriptor is shown in Fig. 3. Assume $\mathbf{L}$ contains n vertices except the keypoint, n^2-2n triangles can be generated based on the KMTS scheme. Consequently, the dimensionality of each feature vector is n^2-2n. Next, we normalize every entry of the feature vector into the range of $[-1, +1]$ and quantize them into histograms by counting the number of entries falling into each of the m bins ($m=30$ in our case). We then concatenate all of the four histograms to form a histogram-based descriptor. Finally, each local patch $\mathbf{L}$ can be represented by the proposed KMTS descriptor, with a length of $m\times 4$. An illustration of the four generated histogram statistics of the proposed KMTS feature collected around a detected keypoint is given in Fig. 4.

4. A TPWCRC based 3D face recognition framework

4.1. Motivation of TPWCRC

SR has become a powerful technique to address many pattern recognition and computer vision problems [36]. SR assumes that each data point $\mathbf{y}\in\mathbb{R}^m$ can be encoded as a sparse linear combination of other points from a dictionary. That is, $\mathbf{y}=\mathbf{Dx}$, where $\mathbf{D}$ is a dictionary of training samples, and $\mathbf{x}$ is the representation coding of $\mathbf{y}$ over $\mathbf{D}$. It is required that most entries in $\mathbf{x}$ are zeros. This can be calculated by solving the following l_1-minimization problem:

$$\hat{\mathbf{x}}=\arg\min_{\mathbf{x}}\|\mathbf{Dx}-\mathbf{y}\|_2^2+\lambda\|\mathbf{x}\|_1 \tag{4}$$

where $\lambda\geq 0$ is a scalar parameter to make a balance between the reconstruction error and the sparsity. However, it is argued that

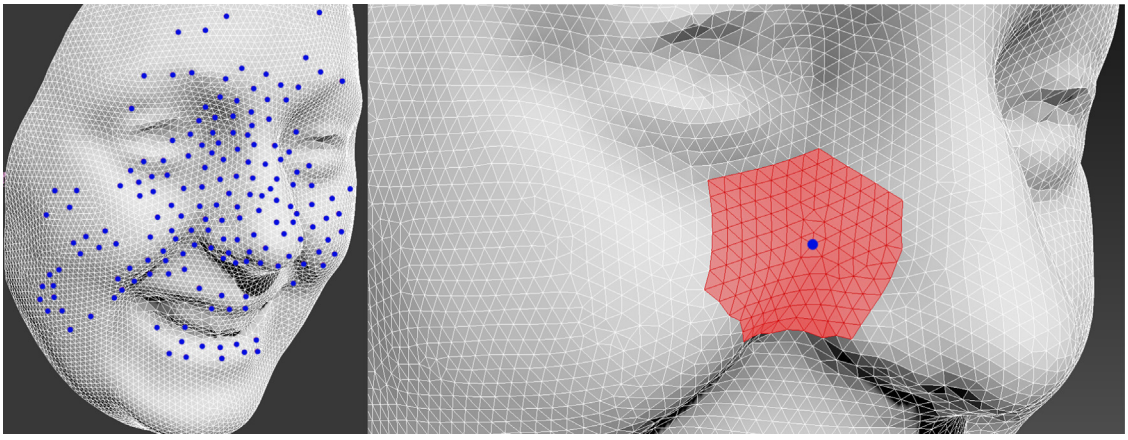

Fig. 2. An illustration of the detected keypoints and the local patch around a keypoint.

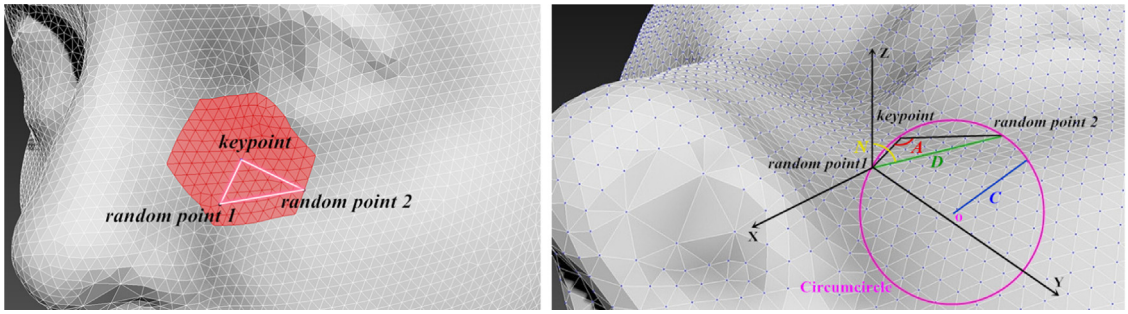

Fig. 3. An illustration of the proposed four types of KMTSs around a detected keypoint on a facial scan.

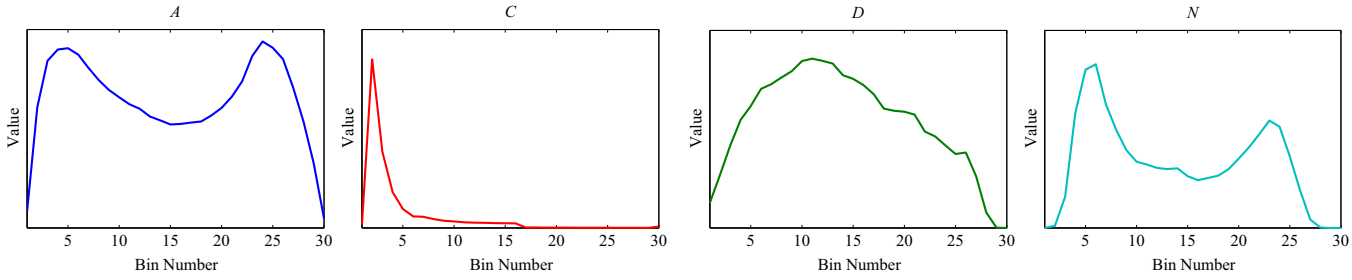

Fig. 4. The four histogram descriptors of the proposed KMTS feature generated for a detected keypoint.

the l_1-norm based sparsity is actually not necessary for FR [13]. The authors argue that the Collaborative Representation (CR) plays a more important role in FR compared to l_1-norm sparsity. As a result, a weaker l_2-norm regularization term is used to replace the strong l_1-norm and a CRC with regularized least square (CRC-RLS) scheme is proposed. Compared to SRC, CRC-RLS can achieve similar classification results with significantly lower computational complexity:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{D}\mathbf{x}-\mathbf{y}\|_2^2+\lambda\|\mathbf{x}\|_2^2 \tag{5}$$

Note that multiple facial scans per individual are used to obtain the dictionary in both CR and SR. In contrast, we aim to address the 3D PFR problem with single sample per individual. We represent each facial scan with a set of local KMTS descriptors. It is empirically known that local descriptors are more sensitive in describing facial shapes and share less similarities compared to holistic descriptors [20]. Moreover, the lack of training samples will significantly affect the discriminative capability of data representation. Therefore, both CR and SR will become unsuitable in this case, and more discriminative information is required to enhance data representation. Consequently, we propose a TPWCRC framework to address the single sample based 3D PFR problem for the following considerations. First, in the scenario of single sample based 3D PFR, no matter l_1 or l_2-norm regularization is used, any test sample will significantly differ from the linear span of the single sample of the same individual. This is caused by the lack of locality information and will result in a large representation error. Therefore, the locality information should be enhanced to prevent the data representation to become global and thus improve the performance of single sample based 3D FR. Second, each facial surface is represented with a set of local KMTS descriptors, a class-based probability can therefore be estimated by simple local distance-based classification and sum pooling for these local descriptors. The probability estimated by local descriptors is more accurate compared to holistic descriptors. Third, each class-based probability estimation can be considered as a weight vector, where each entry of the vector indicates the probability for classifying the test sample to the corresponding individual. This weight vector can further be used as a prior classification knowledge to enhance the locality of data representation, and thus reduce representation error. Last, unlike SRC based algorithms which usually require computationally demanding optimization procedures, our TPWCRC based framework has an analytical solution and can therefore be solved efficiently.

4.2. The first-phase classification

As discussed above, either CR or SR aims to represent each new input as a linear combination of the training samples. These algorithms encode a global relationship between the test sample and the dictionary of existing training samples. If the number of training samples is insufficient, the representation will be global and deviate from the target individual. In this paper, we propose a locality preservation strategy to improve the local consistency in the linear coding scheme. Note that although few element-based locality preservation schemes (e.g., [47]) have been proposed, they may result in large errors because there is only one chance for probability estimation in the case of single sample based 3D PFR. We therefore propose a class-based locality preservation scheme by considering all the KMTS descriptors. Consequently, our class-based locality preservation scheme can provide a more reliable locality information.

For this purpose, a local distance-based classification scheme is proposed to calculate the class-based probability estimation $\mathbf{p}(\mathbf{Y}|\mathbf{D}_c)$ of the input face $\mathbf{Y}$ ($\mathbf{Y}=[\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_n]$, where $\mathbf{y}_i$ denotes the ith KMTS descriptor) over the entire dictionary, where $\mathbf{D}_c$ are the training descriptors for the cth individual. The output of the first-phase classification is $\mathbf{p}=[\mathbf{p}(\mathbf{Y}|\mathbf{D}_1),\ \mathbf{p}(\mathbf{Y}|\mathbf{D}_2),\ldots,\mathbf{p}(\mathbf{Y}|\mathbf{D}_C)]$, which is a vector with C entries, where C is the number of all individuals, and each entry indicates the prior possibility for classifying the test face to the corresponding individual. Specifically, the minimal distance $\mathbf{d}_c$ between the test sample $\mathbf{Y}$ and the cth individual can be calculated as

$$\mathbf{d}_c^i = \min(\|(\mathbf{y}_i,\mathbf{D}_c)\|) \tag{6}$$

$$\mathbf{d}_c = \sum_{i=1}^{n}\mathbf{d}_c^i \tag{7}$$

where $\|\cdot\|$ represents the Euclidean distance, $\mathbf{d}_c^i$ denotes the minimal distance between the ith KMTS descriptor $\mathbf{y}_i$ of the input test and the training descriptors for the cth individual. $\mathbf{D}_c$ is the subdictionary of the cth individual with all the training descriptors, and $\mathbf{d}=[\mathbf{d}_1,\mathbf{d}_2,\ldots,\mathbf{d}_C]$ denotes the minimal distance between $\mathbf{Y}$ and all the C individuals. Finally, a probability vector $\mathbf{p}$ is calculated as

$$\mathbf{p}(\mathbf{Y}|\mathbf{D}_c)=\exp\left(\frac{\mathbf{d}_c/\mathbf{d}_{\min}}{\varepsilon}\right) \tag{8}$$

where $\mathbf{d}_{\min}=\min(\mathbf{d})$, ε is a constant. The proposed probability estimation scheme enhances the locality information when the training samples of each individual is insufficient. Furthermore, it prevents the data representation from becoming global as it imposes a locality constraint in the second-phase of classification.

4.3. The second-phase classification

As described above, a facial scan is represented by a set of KMTS descriptors for both training and test samples. Even for the single-sample problem, the training samples may contain thousands of atoms determined by the number of individuals in the gallery. Furthermore, unlike any holistic descriptor based methods, KMTS lacks similarity even between descriptors captured for the same individual. As a result, according to the characteristics of our KMTS descriptor, we filter the training samples to generate a corresponding gallery dictionary for each of the KMTS descriptor.

We first apply the K-Nearest Neighbor (K-NN) search with Euclidean distance on the training samples of all the C individuals to extract the subdictionary $\mathbf{D}_c=[\mathbf{D}_c^1,\ \mathbf{D}_c^2,\ldots,\mathbf{D}_c^K]$ for the cth individual, where K can be specified as a constant in this work. Therefore, a subset training dictionary can be generated by $\mathbf{D}=[\mathbf{D}_1,\ \mathbf{D}_2,\ldots,\mathbf{D}_C]$, where C denotes the total number of training individuals.

Note that it is argued that the locality is more important than sparsity [47]. This is because the locality will lead to sparsity but sparsity may not result in locality. Therefore, we present a weighted collaborative representation classification scheme to incorporate the weighting term with the learned class-based probability estimation $\mathbf{p}$ (see Section 4.2). Specifically, for each probe KMTS descriptor $\mathbf{y}_i$, the WCRC coding process uses the following objective function:

$$\hat{\mathbf{x}}_i=\arg\min_{\mathbf{x}_i}\|\mathbf{D}\mathbf{x}_i-\mathbf{y}_i\|_2^2+\lambda\|\mathbf{p}^T\mathbf{x}_i\|_2^2 \tag{9}$$

where $\mathbf{x}_i$ is the coding coefficient vector for the ith KMTS descriptor. Unlike the typical SRC and CRC which might select different bases to achieve sparsity, the WCRC selects bases from local neighbors to ensure the correlations between codes. Consequently, similar descriptors can have similar codes. The solution of WCRC can be derived analytically, which avoids the computationally demanding optimization procedures. An illustration of the proposed KMTS–TPWCRC is shown in Fig. 5. It can be observed that incorporating the probability estimation as weight vector (negative polarity) can significantly enhance the reconstruction sparsity in the case of single sample problem and thus improve the recognition performance.

Finally, the class specific representation residual is used to determine the recognition result of the probe test:

$$r_c(\mathbf{Y})=\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{y}_i-\mathbf{D}_c\delta_c(\hat{\mathbf{x}}_i)\|_2/\|\delta_c(\hat{\mathbf{x}}_i)\|_2 \tag{10}$$

where $\delta_c(\cdot)$ is a function which only selects the coefficients corresponding to cth individual. Eq. (10) applies a sum pooling scheme among reconstruction residuals by considering all the n KMTS descriptors with respect to each individual. The overall reconstruction residuals $r=(r_1,r_2,\ldots,r_C)$ can be considered as a similarity score with negative polarity.

For face verification, given a threshold η:

$$\begin{aligned}&\text{Accept}\quad \text{if } \eta-r_c(\mathbf{Y})\geq 0\\ &\text{Reject}\quad \text{if } \eta-r_c(\mathbf{Y})<0.\end{aligned} \tag{11}$$

Both the Verification Rate (VR) and the False Acceptance Rate (FAR) will increase with the threshold η.

For face identification, the identity $I(\mathbf{Y})$ of a test $\mathbf{Y}$ can be computed as

$$I(\mathbf{Y})=\arg\max_{1\leq c\leq C}(r_c(\mathbf{Y})), \tag{12}$$

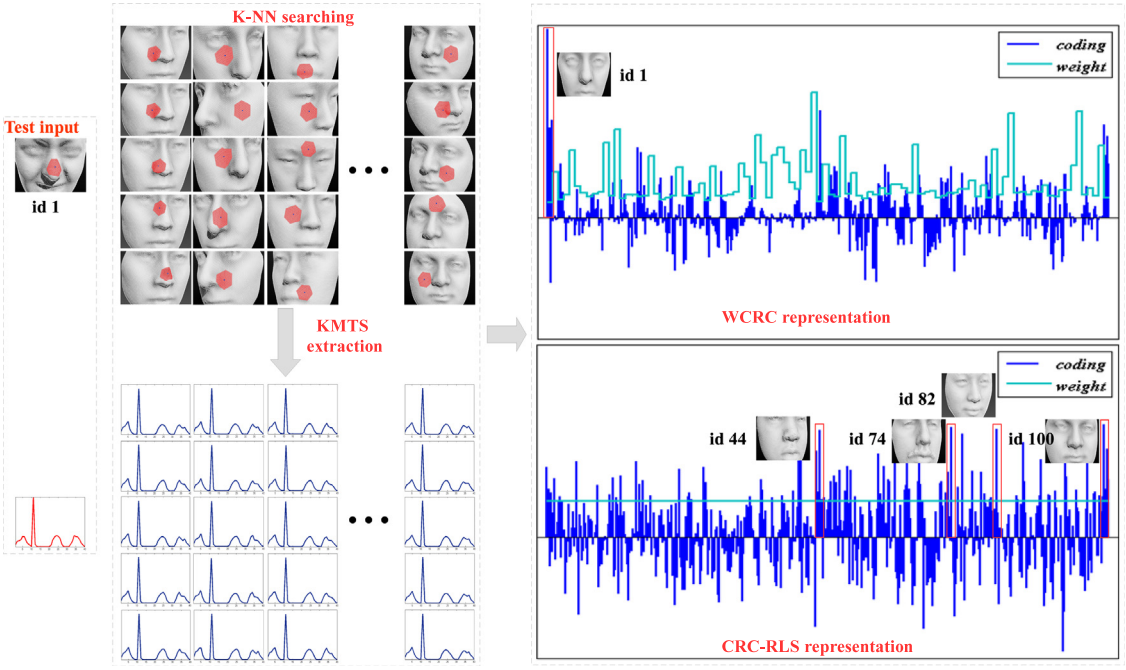

Fig. 5. An illustration of the proposed KMTS–TPWCRC framework for 3D PFR. It can be seen that the incorporation of the prior classification knowledge significantly concentrates the coding to some specific individuals and thus improves the recognition performance.

Fig. 6. Examples of the preprocessing results. The raw facial scans contain some obvious missing data, spikes and undesired parts.

where the identity is determined by the individual that yields the largest $r_c(\mathbf{Y})$.

5. Experimental results

We tested our proposed KMTS-TPWCRC 3D PFR framework on six challenging 3D face datasets including Bosphorus, GavabDB, UMB-DB, SHREC 2008, BU-3DFE and FRGC v2.0.

In this section, we first introduce our 3D data preprocessing in Section 5.1. We then present the results of our proposed approach on partial facial datasets in Section 5.2. Comparative results of our approach with respect to large expression deformations are presented in Section 5.3. Comparative results of different classifiers and the parameter analysis are provided in Section 5.4. Computational complexity of the proposed approach is analyzed in Section 5.5.

5.1. Data preprocessing

Most of the raw facial data used in our experiments suffers from a number of nuisances (including noise spikes, holes, pose variations, and occlusions by parts such as clothes, neck and ears). Therefore, suitable data preprocessing should be performed before face recognition. Our four-step facial data preprocessing algorithm is developed based on the Visualization Toolkit (VTK), which is an open-source software for 3D computer graphics, image processing and visualization.[1] These steps are explained below.

First, the nosetip is automatically detected on a facial surface using the algorithm proposed in our previous work [48,49]. We define a sphere with a radius of 80 mm centered at the nosetip to crop the 3D face by discarding the points outside the sphere. Second, the noise spikes are reduced by eliminating the outlier vertices, resulting in a smoothed facial mesh. The missing parts are automatically identified by locating boundary edges, iteratively linking them into loops, and performing triangulation on the obtained loops. Third, we apply our pose correction algorithm [48] to correct the pose of a 3D facial scan to its frontal view. Finally, we resample the facial mesh onto a square grid of 161×161 mm with a uniform resolution of 1 mm. Fig. 6 shows some examples of the preprocessed facial scans.

The proposed preprocessing pipeline is only applied for the facial scans in the Bosphorus, GavabDB, UMB-DB datasets. This is because the facial scans of these datasets have poor quality and suffer from data corruption and noise. Most of the existing approaches perform sophisticated preprocessing steps for their experiments on these datasets. For a fair comparison of our experimental results with these approaches, we apply the proposed automatic preprocessing pipeline (i.e., nosetip detection, face cropping, mesh enhancement and resampling) to normalize the facial scans in these three datasets only. For some of the facial scans from GavabDB with large head rotations and extreme profile views, automatic nosetip detection fails. Following a similar experimental setup as previous works [15], we manually detect the nosetip for these extreme cases in GavabDB. Apart from these extreme scans with profile views, the proposed approach is fully automatic provided the face can correctly be cropped. Under extreme cases when the nosetip is invisible, manual intervention or other more advanced face detection and cropping algorithm is required.

For the experiments performed on SHREC 2008, BU-3DFE and FRGC v2.0 datasets, we only involve simple preprocessing steps. Specifically, since these datasets are relatively high quality and most of the previous approaches evaluated on these datasets have only performed some simple preprocessing (e.g., spikes removal and hole filing) without any further mesh enhancement. For the consideration of a fair comparison with these methods, we only automatically detect the nosetip, crop the facial region and remove spikes/fill holes without any mesh enhancement process as was done for the Bosphorus, GavabDB, UMB-DB datasets.

5.2. 3D face recognition under missing parts, occlusions and data corruptions

5.2.1. Results on the Bosphorus dataset

The Bosphorus dataset [28] contains 4652 facial scans belonging to 105 individuals (60 men and 45 women aged between 25 and 35). The number of scans is more than 31 for each individual. These scans have been recorded under different poses, expressions and external occlusions. These occlusions include: (1) occlusion of the mouth with hand, (2) glasses, (3) occlusion of the face with hair, and (4) occlusion of the left eye and forehead regions by

[1] The codes are publicly available at their website: www.vtk.org.

hands, as illustrated in Fig. 7. Before the release of Bosphorus dataset, only a limited number of 3D face datasets with external occlusions were publicly available. The Bosphorus dataset therefore provides a suitable and large benchmark for the evaluation of 3D PFR algorithms.

Only a few existing approaches have addressed the 3D PFR problem, e.g., [17,31,10,15]. To achieve rigorous and fair comparison, we follow the same evaluation protocol as described in [15]. Specifically, one scan with neutral expression is selected for each individual to generate a gallery. Then, 381 facial scans with external occlusions (i.e., eye, mouth, glasses and hair) are selected to form the probe set. The rank-1 Identification Rate (IR) results of our proposed approach are presented in Fig. 8(a), with comparison to the state-of-the-arts (as shown in Table 1(a)). Our approach achieves an overall rank-1 IR of 92.7%.

On the Bosphorus dataset, the approach proposed by Li et al. [16] achieves the highest overall rank-1 IR of 99.2%. Their method uses two sophisticated keypoint detectors and three keypoint descriptors to encode a facial scan. The major limitation of the method is that it is computationally expensive and not suitable for real-world applications. Compared with [16], our approach employs only simple local features and is therefore computationally efficient. Specifically, in the facial feature extraction phase, the proposed approach is almost 10 times faster than the approach proposed by Li et al. [16], and during the classification phase, our approach is 2 times faster than their approach. We believe that the recognition performance of the proposed approach is subject to the feature extraction algorithm. That is, if more sophisticated features are incorporated in the proposed approach to represent a facial scan, the recognition performance can be significantly improved. However, for the consideration of the practical applications, we propose a framework which is capable of handling the 3D PFR in the presence of single training sample and achieves reasonable performance while being more suitable for real-world applications.

Colombo et al. [10] reduced the size of probe set to 360 by removing 21 facial scans with extremely poor quality (see some examples in Fig. 9). We follow their protocol and report the recognition results on the reduced probe set in Fig. 8(b). It can be observed that the proposed approach obtains the highest overall rank-1 IR compared to the other two approaches, as shown in Table 1(b). Although our approach performs slightly worse in the case of eye occlusion, we achieve superior performance for the other three cases (i.e., mouth, glasses and hair occlusions). These results fully illustrate the robustness of the proposed approach under partial facial data caused by external occlusions.

In addition to occlusions, the Bosphorus dataset also contains a large number of expressions and pose variations. To evaluate the performance of the proposed approach for this aspect, we performed another set of experiments following the same experimental setup as [16]. Specifically, the gallery is defined as the first neutral scan for each individual. Then, two sets of probes are selected. The first set (called the expression subset) contains 6 basic facial expressions (i.e., anger, disgust, fear happiness, sadness and surprise), and 28 facial Action Units (i.e., 20 Lower AUs, 5 Upper AUs, and 3 Combined AUs). The second probe set (called the pose subset) includes 7 Yaw Rotations (i.e., $+10°$, $+20°$, $+30°$, $\pm 45°$, and $\pm 90°$), 4 Pitch Rotations (i.e., strong upwards/

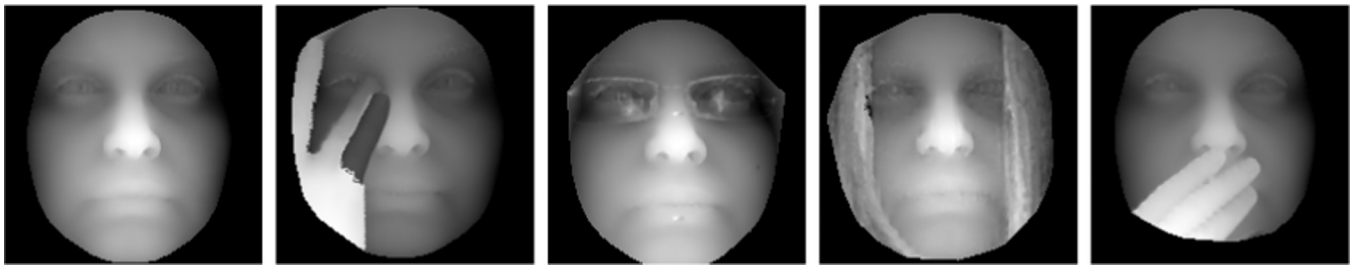

Fig. 7. Examples of different occlusions of the facial scans present in the Bosphorus dataset. The unoccluded facial scan is shown on the left, followed by four types of occlusions (eye, glasses, hair and mouth).

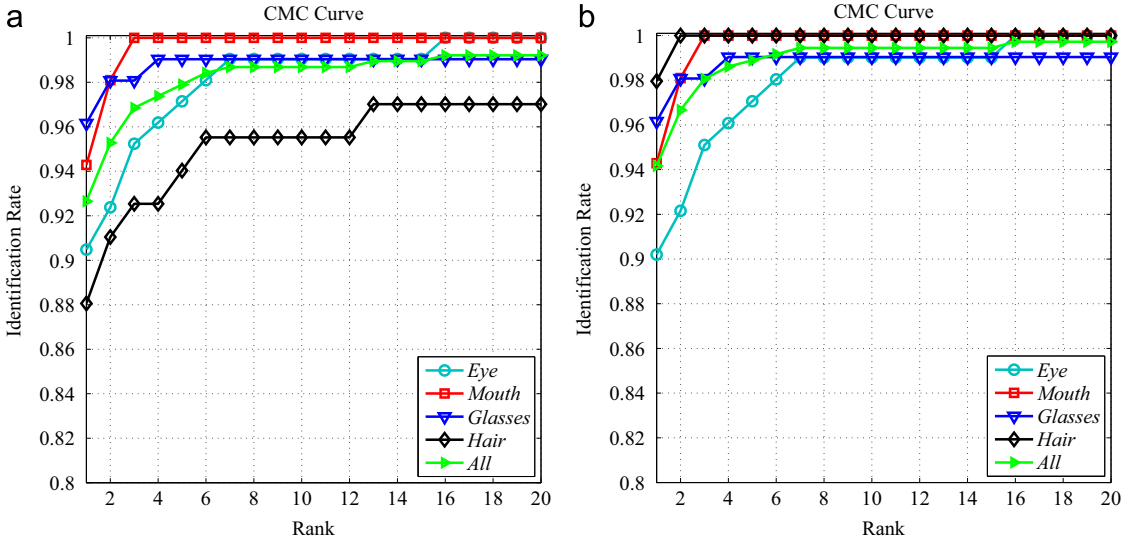

Fig. 8. The CMC curves of the face identification experiments with different occlusions on the Bosphorus dataset. (a) The results using all the probe scans and (b) the results using the selected 360 probe scans which exclude the scans with extremely poor quality.

downwards and slight upwards /downwards) and 2 Cross Rotations (i.e., +45° yaw and ±20 pitch). The achieved experimental results in terms of rank-1 IRs and the comparisons with the state-of-the-art approaches are given in Table 2.

From Table 2, it can be observed that the proposed approach achieves the highest performance on the subset of facial expressions with a rank-1 IR of 98.9%, and a competitive performance compared to the best reported results in [16] on the subset of facial poses (rank-1 IRs of 90.6% vs. 91.1%). The proposed approach outperforms all the other approaches (i.e., [50–53]) on all of the subsets. It should be noted that the approach in [16] relies on a complicated and time-consuming feature extraction algorithm and the other compared approaches heavily rely on complicated registration algorithms. In contrast, the proposed approach incorporates only some low-level geometrical features which do not require any complicated mathematical operations. These results indicate that the proposed approach can achieve a similar performance under facial expression and pose variation on the Bosphorus dataset and is more computationally efficient (see Section 5.5).

Note that the proposed approach achieves the best performance (a rank-1 IR of 47.4%) for faces with large pose variations (i.e., YR 90°) since our low-level geometrical features are directly computed from the spatial relationships of the 3D facial vertices. Face pose variations and other rigid motions can only change the absolute spatial positions of the 3D vertices on a facial surface. However, the relative local spatial relationships among those vertices remains unaffected, which enables our approach to work effectively on nearly profile facial scans.

5.2.2. Results on GavabDB

GavabDB [9] contains 3D facial scans exhibiting a wide range of variations caused by missing parts, occlusions and data corruptions. The GavabDB dataset was captured with a Minolta Vivid sensor and contains facial scans of 61 Caucasian individuals (45 male and 16 female). Each individual was scanned from different angles and under various expressions including six neutral and three nonneutral expressions (i.e., smile, laugh, and an arbitrary expression). Moreover, some neutral scans are with large pose variations, e.g., one looking up scan (+35°), one looking down scan (−35°), one left profile scan (−90°) and one right profile scan (+90°). Therefore, partial data, large pose variations and facial expressions make GavabDB very challenging for 3D PFR. We follow the same evaluation protocol as [54], i.e., the first frontal scan with neutral expression of each individual is used to form the gallery (61 in total), and the remaining scans are used to form the probe set.

Table 3 (a) presents a comparison of the achieved experimental performance of different methods on GavabDB. It can be observed that the proposed approach outperforms most the of the existing approaches. Specifically, the proposed approach achieves the highest rank-1 IR of 95.08% for the probe scans with nonneutral expressions. It also achieves the highest rank-1 IR of 96.31% for probe scans with both neutral and nonneutral expressions. Analytically, it is more challenging to recognize scans with large pose variations since more parts are occluded in these scans (looking up and looking down). Nevertheless, the proposed approach still achieves a very high rank-1 IRs of 98.36% for the probes with pose variations up to 35° (both looking up/down) (Table 4).

In order to further evaluate the robustness of the proposed approach with respect to the sideway scans on GavabDB, we conducted an experiment following the protocol introduced by previous publications [15,58,53]. Table 3(b) presents the comparative results tested on those sideway scans. Since only half of the facial region is visible for these scans, we observe a decrease in the achieved performance. Our approach outperforms [15], with rank-1 IRs of 70.49% and 86.89% for the scans from right and left, respectively. Note that two local feature matching-based approaches (i.e., [58,53]) achieved higher recognition performance on sideway scans. That is, the local feature matching-based approaches can incorporate the spatial relationship between the local features and thus can improve the matching accuracy. However,

Table 1
Recognition performance under different occlusions on the Bosphorus dataset. (a) The rank-1 IRs on all the 381 occluded probe scans and (b) the rank-1 IRs on the selected 360 probe scans excluding the scans with extremely poor quality.

Approaches	Eye (%)	Mouth (%)	Glasses (%)	Hair (%)	Overall (%)
(a)					
Alyuz et al. [28][a]	93.6	93.6	97.8	89.6	93.6
Colombo et al. [10]	91.1	74.7	94.2	90.4	87.6
Drira et al. [15]	97.1	78	94.2	81	87
Li et al. [16]	**100**	**100**	**100**	95.5	**99.2**
Ours	90.5	94.3	96.2	88.1	92.7
(b)					
Colombo et al. [10]	91.1	74.7	94.2	90.4	87.6
Drira et al. [15]	**98.9**	78.5	94.2	85.2	89.2
Ours	90.2	**94.3**	**96.2**	**98**	**94.2**

[a] The approach is not automatic and requires manual intervention.

Table 2
Comparison of rank-1 IR results on the subsets of facial expressions and poses on the Bosphorus dataset.

Approaches	Exp (%)	yr	yr 90°	pr	cr	pos
Alyuz et al. [50]	98.2	–	–	–	–	–
Ocegueda et al. [51]	98.2	–	–	–	–	–
Smeets et al. [52]	97.7	–	24.3%	–	–	84.2%
Berretti et al. [53]	95.7	81.6%	45.7%	98.3%	93.4%	88.6%
Li et al. [16]	98.8	84.1%	47.1%	**99.5%**	**99.1%**	**91.1%**
Ours	**98.9**	83.8%	**47.4%**	98.3%	98.6%	90.6%

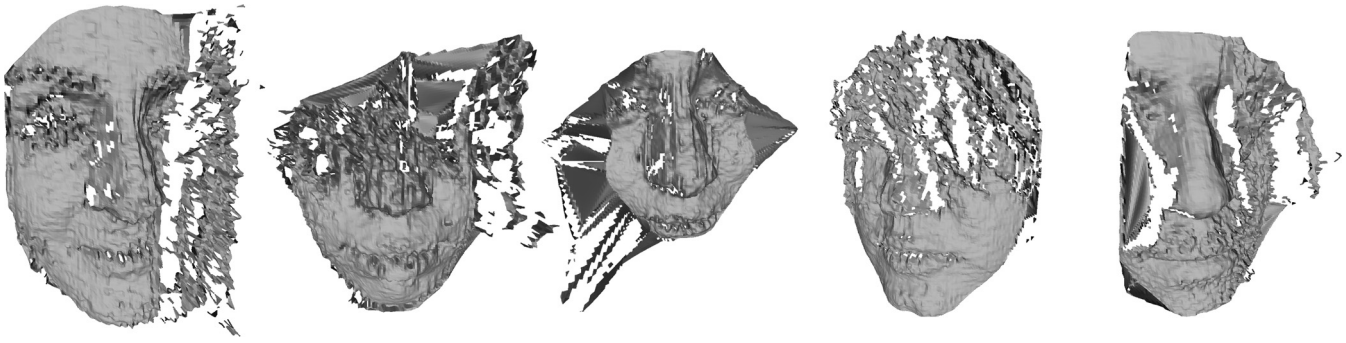

Fig. 9. Examples of the eliminated facial scans from the occlusion subset of the Bosphorus dataset with extremely poor quality.

Table 3
Recognition performance on GavabDB. (a) The rank-1 IRs on different subsets and (b) the rank-1 IRs on the sideway scans.

Approaches	*n*	*e*	*n*+*e*	*u*	*d*	*o*
(a)						
Li et al. [24]	96.67%	93.33%	94.68%	–	–	–
Moreno et al.[55]	90.16%	77.9%	–	–	–	–
Mahoor et al. [56]	–	72%	78%	85.3%	88.6%	–
Mousavi et al. [57]	–	–	91%	–	–	81.67%
Zhang et al. [40]	–	–	–	–	–	92.26%
Huang et al. [58]	100%	94%	95.5%	96.7%	96.7%	–
Berretti et al. [53]	**100**%	94%	95.1%	96.7%	95.1%	–
Drira et al. [15]	**100**%	94.54%	95.9%	**100**%	**98.36**%	**96.99**%
Ours	**100**%	**95.08**%	**96.31**%	98.36%	**98.36**%	**96.99**%
Approaches			*r* (%)			*l* (%)
(b)						
Drira et al. [15]			70.49			86.89
Huang et al. [58]			78.7			**93.4**
Berretti et al. [53]			**83.6**			**93.4**
Ours			73.77			90.16

n, neutral; *e*, expression; *u*, looking up; *d*, looking down; *o*, overall scans; *r*, sideway scan from right; and *l*, sideway scan from left.

Table 4
Recognition performance on UMB-DB in terms of rank-1 IR for the "neutral vs. occlusion" experiment.

Approaches	*n* vs. *o* (%)
UMB-DB baseline	56.5
Alyuz et al. [17][a]	70.51
Ours	**73.08**

[a] The automatic occlusion masking method is selected for comparison.

these approaches are computational expensive and thus not unsuitable for practical applications. Nevertheless, the proposed approach achieves a rank-1 IRs of 73.77% and 90.16% for the sideway scans acquired from right and left sides respectively, which are comparable to the best reported results.

5.2.3. Results on UMB-DB

UMB-DB [10] contains a total of 1473 scans (590 occluded and 883 unoccluded) of 143 individual (98 males and 45 females) with number of scans for each individual ranging from 7 and more). Notice that, the facial scans in UMB-DB are considered to be more challenging in terms of occlusions. These occlusions can be caused by hair, hands, hair scarves, eyeglasses and other objects (e.g., scissors), as shown in Fig. 12. Additionally, the location and the content of these occlusions vary more significantly as compared to the Bosphorus and GavabDB datasets. Meanwhile, the facial expression variations presented in the scans from UMB-DB are relatively not rich as other 3D face datasets (e.g., FRGC v2.0 and BU-3DFE datasets). Therefore, UMB-DB is more suitable to evaluate the 3D PFR approaches.

Following the same experimental protocol as [10]. We select the first neutral scan of each individual to form the gallery set (143 in total), and the probe set is consisted of all the 590 occluded scans. The CMC curve of the proposed 3D PFR approach for the aforementioned "neutral vs. occlusion" experiment is presented in Fig. 10, and the corresponding comparison between the proposed approach and the state-of-the-art results is presented in Table 1 (a). The proposed approach achieves a rank-1 IR of 73.08%, which outperforms the reported result of 70.51% in [17]. Notice that their approach includes manual/automatic methods for occlusion masking, so we only compare our approach with their automatic occlusion masking method.

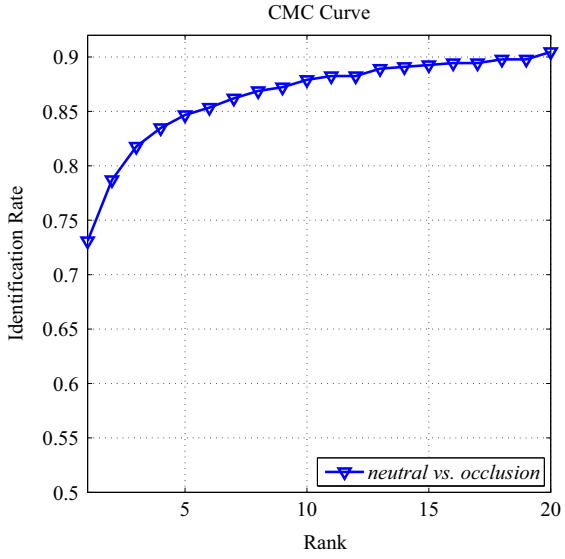

Fig. 10. The CMC curve of the proposed approach for the "neutral vs. occlusion" experiment on UMB-DB.

Some existing approaches requires a sophisticated occlusion detection and restoration process to handle 3D PFR (e.g., [17]). Since, the occlusion variation of a facial scan is complicated in real-world applications, a occlusion detection and restoration algorithm may fail and consequently affect the face recognition performance. Moreover, the existing occlusion detection and restoration methods mainly rely on the ICP-based algorithm or one of its variants, which is computationally expensive. However, our approach does not require occlusion detection/restoration and is thus computationally efficient for real-world applications.

5.2.4. Performance evaluation for robustness against partial facial data on the Bosphorus dataset, GavabDB and UMB-DB

In order to evaluate the performance of the proposed approach in presence of missing parts, occlusions and data corruptions, extensive experiments have been performed on the Bosphorus, GavabDB and UMB-DB datasets. The proposed approach obtained competitive results as compared to the state-of-the-arts. Several conclusions can be drawn from the experimental results.

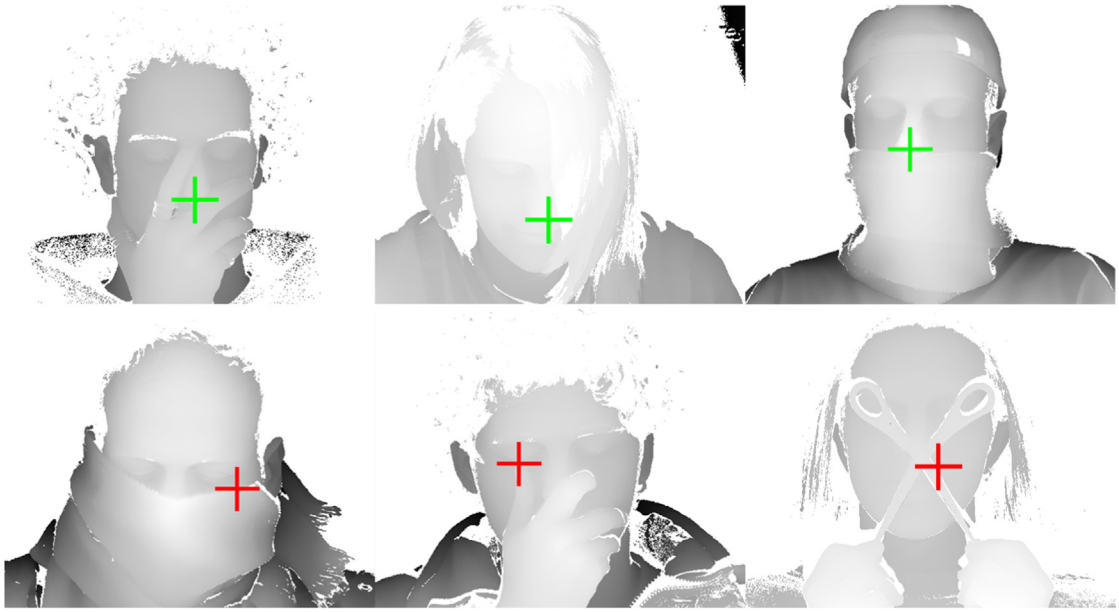

Fig. 11. Some nose detection results from UMB-DB. Note that the correct detections are given in the first row and the incorrect ones are given in the second row.

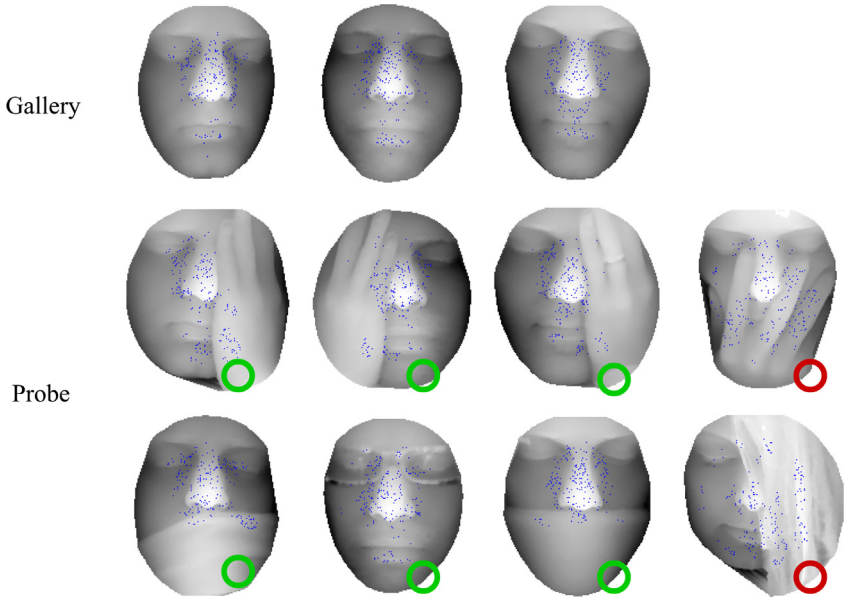

Fig. 12. An illustration of keypoint detection and face recognition results for the difficult facial scans (i.e., external occlusions). Our detected keypoints mainly located on the nose/eye–forehead area. Our approach may fail only then the majority part of the nose/eye–forehead area is occluded. Note that a green circle means correct recognition and a red circle means incorrect recognition. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

First, the proposed approach represents a facial scan using a set of the KMTS descriptors. Compared to holistic descriptors, the proposed KMTS descriptor is more suitable for partial 3D face recognition. Second, unlike some existing 3D PFR approaches (e.g., [15,17]), the proposed approach does not require any complicated data restoration process (which is typically based on the ICP algorithm or its variants) to recognize facial scans with data corruptions. Third, the proposed approach relies on a KMTS–TPWCRC framework to efficiently handle the 3D PFR problem. For this purpose, it first applies an automatic nosetip detection algorithm and a 3D facial keypoint detection algorithm. The two algorithms are only based on simple computations and give promising results. Our nosetip detection is able to achieve around 99% detection accuracy for those facial scans with nose area fully visible. Fourth, for the highly occluded facial scans (e.g., especially from UMB-DB), our nosetip detection algorithm can still achieve a 94% accuracy (see some of the correct/incorrect detection results in Fig. 11). The erroneous nosetip detection will lead to an inaccurate face cropping. Nevertheless, our approach uses a simple and efficient 3D facial keypoint detection algorithm to extract useful salient features for 3D facial scan representation. Although inaccurate face may be cropped (the center of cropped face is not located in the nosetip), our coordinate-independent keypoint detection algorithm can still extract enough discriminative information. Lastly, the extracted keypoints are mainly located on the nose and eye-forehead area, these semi-rigid area was proven to be more discriminative and useful for face recognition in our previous work [11]. For those occluded facial scans with only nose and eye-forehead areas visible (see the first three columns of Fig. 12), our approach can still achieve correct classification. The proposed approach only fails on faces with major portion of the nose and eye-forehead areas

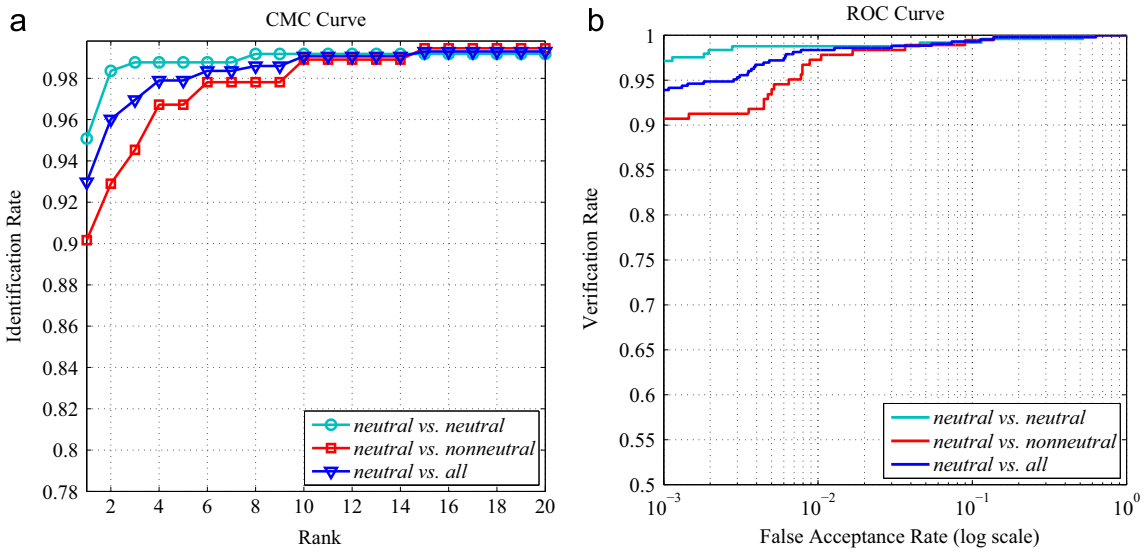

Fig. 13. The CMC and ROC curves on the SHREC 2008 dataset. (a) The CMC curves and (b) the ROC curves. The resulting rank-1 IR and 0.1% FAR VR of the "neutral vs. all" experiment are 93% and 93.91%, respectively.

Table 5
Comparative results in terms of 0.1% FAR VRs on the SHREC 2008 dataset.

Approaches	*n* vs. *n* (%)	*n* vs. *nn* (%)	*n* vs. *a* (%)
Berretti et al. [26]	96.31	80.87	90.40
Amberg et al. [59]	96.72	74.86	88.52
ter Haar et al. [60]	70.90	39.34	55.50
Nair et al. [61]	60.25	42.62	52.93
Xu et al. [62]	47.95	43.17	44.96
Ours	**97.13**	**90.71**	**93.91**

n vs. *n*, neutral vs. neutral; *n* vs. *nn*, neutral vs. nonneutral; and *n* vs. *a*, neutral vs. all.

occluded (i.e., the hair occlusion on the Bosphorus dataset, see the last column of Fig. 12).

5.3. 3D face recognition under large expressions

5.3.1. Results on the SHREC 2008 dataset

The SHREC 2008 dataset comprises 427 3D facial scans belonging to 61 individuals (45 males and 16 females). Seven scans were acquired for each individual with various facial expressions. Specifically, two frontal scans are with neutral expressions, another two neutral scans were with slightly rotated postures (looking up/looking down), and the remaining three scans were with nonneutral expressions (smile, laugh or an arbitrary expression). In summary, the SHREC 2008 dataset consists of 244 scans with neutral expression and 183 scans with non-neutral expressions.

Following the same evaluation protocol as [26], one neutral scan of each individual is selected to form the gallery set (61 in total). Three experiments are performed for performance evaluation and comparison: (1) two neutral scans with looking up/down postures are included in the probe set, namely "neutral vs. neutral" experiment; (2) three nonneutral scans are used to form the probe set, namely "neutral vs. nonneutral" experiment; and (3) both of the neutral and nonneutral scans are used for evaluation, namely the "neutral vs. all" experiment.

The CMC and ROC curves of our proposed approach on the three experiments are presented in Fig. 13. It can be observed that, the rank-1 IR and the 0.1% FAR VR for the "neutral vs. all" experiment are 93% and 93.91%, respectively. The comparison between our proposed approach and the state-of-the-arts is shown in Table 5.

In all of the three experiments, our proposed approach outperforms all the other approaches. Specifically, our approach slightly outperforms the approach proposed in [59] for the "neutral vs. neutral" experiment, with 0.1% FAR VRs of 97.13% and 96.72%, respectively. Nevertheless, our proposed approach achieves a 0.1% FAR VR of 90.71% in the more challenging "neutral vs. nonneutral" experiment, while the highest rate reported in the literature is only 80.87%. Our approach also achieves the best performance in the "neutral vs. all" experiment and outperforms [26] by a margin of 3.51%. These results clearly demonstrate the robustness of our proposed approach under deformations caused by large facial expressions.

5.3.2. Results on the BU-3DFE dataset

The BU-3DFE dataset contains 2500 textured facial scans of 100 individuals (44 males and 56 females) with a large range of diversity in age and ethnic/racial ancestries. 25 ear-to-ear facial scans were acquired for each individual. One is with neutral expression while the remaining 24 scans are with various expressions (e.g., happiness, anger, fear, disgust, sadness and surprise). Moreover, each expression has four intensities (where levels 1/2 and 3/4 are considered as low and high intensities, respectively). Compared to other 3D face datasets, the BU-3DFE dataset is richer in expressions, and it is therefore quite challenging for 3D FR.

The neutral scan of each individual is selected to form the gallery set, and three different experiments are devised. First, all the low-intensity scans (level 1/2) are selected to form the probe set (1200 faces in total), namely "neutral vs. low-intensity" experiment. Second, all the high-intensity facial scans (level 3/4) are selected to form a more challenging probe set (1200 faces in total), namely "neutral vs. high-intensity". Third, facial scans with all expression intensities (2400 faces in total) are used to form the probe set, namely "neutral vs. all" experiment. The CMC and ROC curves of our approach on the BU-3DFE dataset are shown in Fig. 14. Our approach achieved the rank-1 IRs of 96.83%, 89.67% and 93.25% in the "neutral vs. low-intensity", "neutral vs. high-intensity" and "neutral vs. all" experiments, respectively. The 0.1% FAR VRs of these three experiments are 97.42%, 90.25% and 94%, respectively. The performance degradation from "neutral vs. low-

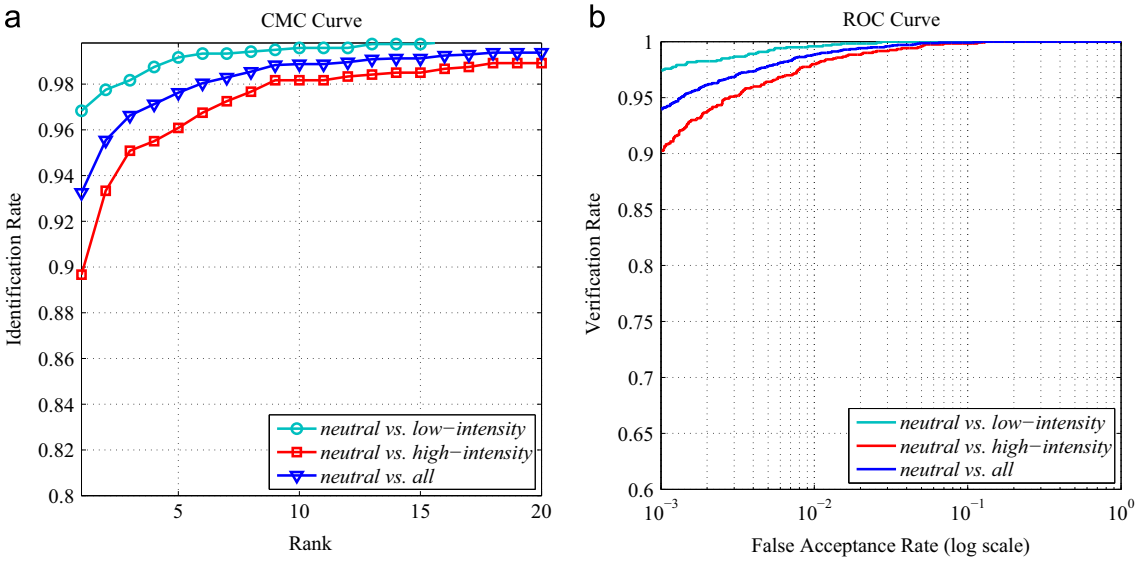

Fig. 14. The CMC and ROC curves on the BU-3DFE dataset. (a) The CMC curves and (b) the ROC curves. The resulting rank-1 IR and 0.1% FAR VR of the "neutral vs. all" experiment on the BU-3DFE dataset are 93.25% and 94%, respectively.

Table 6
Comparison with the state-of-the-art approaches on the BU-3DFE dataset in terms of 0.1% FAR VRs for the "neutral vs. all" experiment.

Approaches	*n* vs. *a* (%)
Ocegueda et al. [63]	**96.3**
Berrtti et al. [53]	87.5
Elaiwat et al. [64][a]	81.5
Ours	94

[a] Only the results on the 3D modal are shown for the corresponding approaches.

intensity" to "neutral vs. high-intensity" is only 7.17% in terms of the 0.1% FAR VR. These results clearly demonstrate the robustness of the proposed approach with respect to large facial expressions.

A performance comparison between our approach and the state-of-the-art approaches is given in Table 6. Our recognition performance (a 0.1% FAR VR of 94%) is comparable to the best reported performance by [63], with a score of 96.3%. However, their proposed approach relies on the assumption that the facial scans are already accurately registered. Thus, this condition limits the practical use of their approach to only well registered facial scans. In contrast, our proposed approach is registration-free and is therefore more robust to complicated variations (e.g., expressions, pose and occlusion variations) for real-world applications. Moreover, their approach requires a time-consuming training process (around 4 h for the BU-3DFE dataset). Nevertheless, our approach is training-free and more suitable for real-world applications since the computational cost during the training phase is also an important concern for an efficient and practical 3D FR system, where the gallery needs to be frequently updated with new enrollments.

5.3.3. Results on the FRGC v2.0 dataset

The FRGC v2.0 dataset is one of the largest publicly available 2D and 3D face dataset [65]. It contains 4007 3D facial scans of 466 individuals. 2410 facial scans are with neutral expression and the remaining 1597 scans are with various nonneutral expressions (e.g., surprise, happy, puffy cheeks and anger). In order to test our proposed approach under the deformations caused by facial expressions, we select the first neutral scan of each individual to form the gallery set (466 in total). Then, three experimental subsets are constructed: (1) "neutral vs. neutral" experiment (1944 probes); (2) "neutral vs. nonneutral" experiment (1,597 probes); (3) "neutral vs. all" experiment (3541 probes). Fig. 16 presents the CMC and ROC curves of our approach for these three experiments. Note that our approach obtains an overall rank-1 IR of 96.3% and a 0.1% FAR VR of 98.3% in the "neutral vs. all" experiment. In the "neutral vs. nonneutral" experiment, our approach achieves a rank-1 IR of 92.2% and a 0.1% FAR VR of 96%, respectively. A rank-1 IR of 99.6% and a 0.1% FAR VR of 99.9% are achieved respectively for the simpler case of "neutral vs. neutral" experiment (Fig. 15).

A performance comparison between our approach and the existing approaches is given in Table 7. Note that we only include the approaches which are specifically developed for the FRGC v2.0 dataset. However, since all of the facial scans in the FRGC v2.0 dataset are in nearly frontal views with very high quality compared to other 3D face datasets (e.g., the Bosphorus and GavabDB), it is not difficult to achieve a good recognition performance by most of the existing 3D FR approaches. However, only a limited existing approaches have explicitly addressed the 3D PFR problem under missing data, occlusions and data corruptions. Besides the ability to deal with the deformations caused by facial expressions (provided by FRGC v2.0), our approach is also able to address the 3D PFR problem (not provided by FRGC v2.0, which is comprehensively tested in Section 5.2). Nevertheless, our approach achieves a very competitive performance (a rank-1 IR of 96.3% and a 0.1% FAR VR of 98.3%) in the "neutral vs. all" experiment on the FRGC v2.0 dataset. Our results are comparable to the best results reported in the literature (a rank-1 IR of 98.4% and a 0.1% FAR VR of 98.6% both reported by [5]).

Open mouth can be considered as a large facial expression since an open mouth significantly deforms a facial surface.[2] In order to evaluate the robustness of our approach with respect to large facial expressions, we perform another evaluation that was first proposed by Berretti et al. [26]. Specifically, the first neutral scan of each individual is selected to form the gallery, and all the facial scans with an open mouth from the FRGC v2.0 dataset are selected as the probes. The resulting CMC and ROC curves for this experiment are shown in Fig. 16. The rank-1 IR and the 0.1% FAR VR achieved by our approach are 89.2% and 93.8%, respectively.

[2] All 816 facial scans were manually selected by Berretti et al. [26], and a list is available at http://www.dsi.unifi.it/~berretti/frgcv2.0/.

Fig. 15. Examples of the facial scans with large mouth deformations from the FRGC v2.0 dataset.

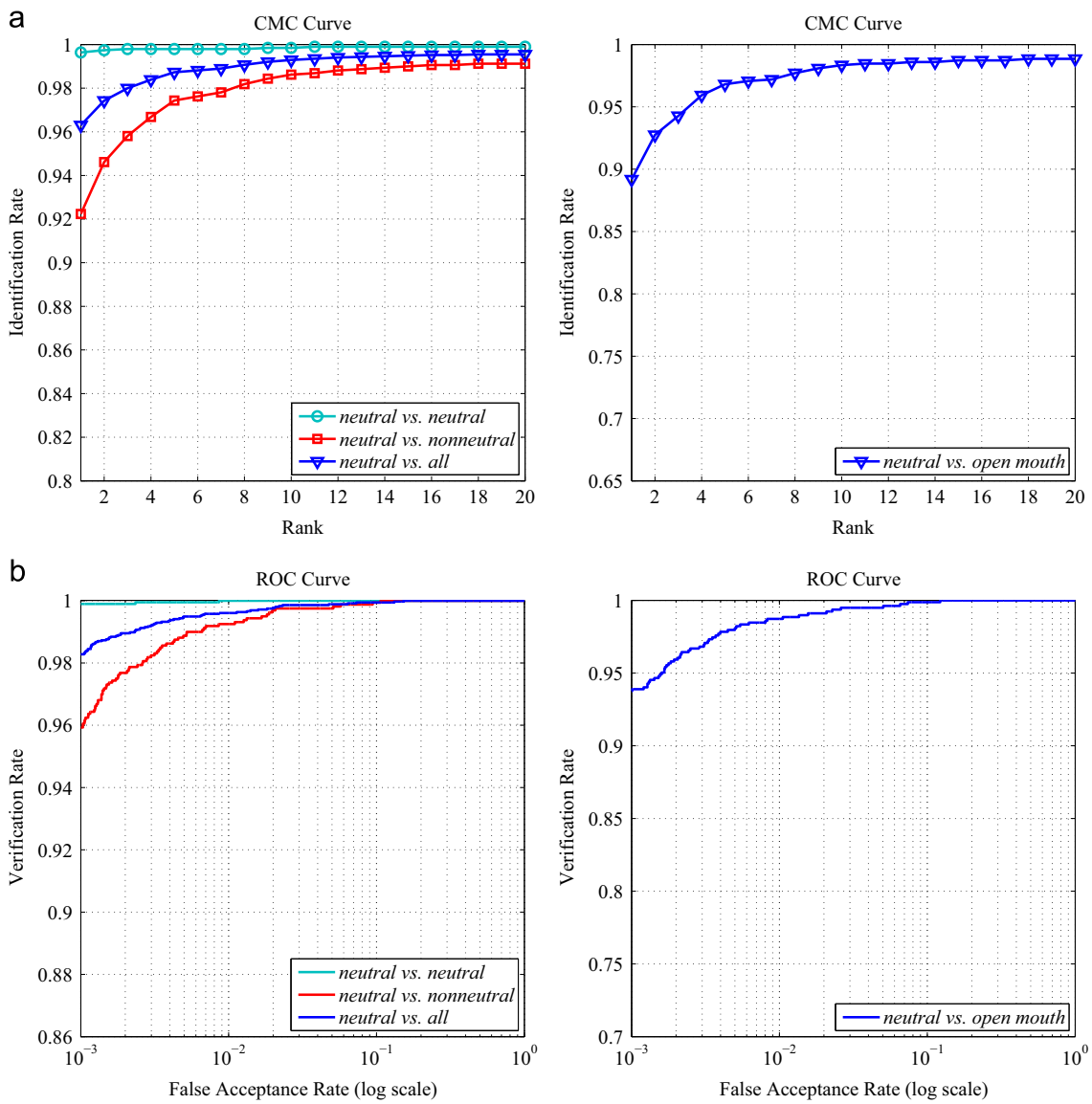

Fig. 16. The CMC and ROC curves for the "neutral vs. neutral", "neutral vs. nonneutral", "neutral vs. all" and "neutral vs. open mouth" experiments on the FRGC v2.0 dataset. (a) the CMC curves and (b) the ROC curves. The resulting rank-1 IR and 0.1% FAR VR of the "neutral vs. all" experiment are 96.3% and 98.3%, respectively. The obtained rank-1 IR and 0.1% FAR VR of the "neutral vs. open mouth" experiment are 89.2% and 93.8%, respectively.

Table 7
Comparison with the state-of-the-art on the FRGC v2.0 dataset. (a) 0.1% FAR VRs for the "neutral vs. neutral", "neutral vs. nonneutral" and "neutral vs. all" experiments and (b) the rank-1 IDs for the "neutral vs. all" experiment.

Approaches	*n* vs. *n*	*n* vs. *nn*	*n* vs. *a*
(a)			
Maurer et al. [66][a]	97.8%	–	86.5%
Mian et al. [48][a]	99.4%	97%	98.5%
Mian et al. [20][a]	**99.9%**	92.7%	97.4%
Al-Osaimi et al. [67]	98.4%	**97.8%**	98.1%
Elaiwat et al. [64][a]	99.6%	93.1%	–
Berretti et al. [26]	97.7%	91.4%	95.5%
Passalis et al. [68]	94.9%	79.4%	85.1%
Wang et al. [5]	–	–	**98.6%**
Huang et al. [58]	99.6%	97.2%	98.4%
Ours	**99.9%**	96%	98.3%
Approaches			*n vs. a* (%)
(b)			
Cook et al. [69]			94.6
Mian et al. [48][a]			96.2
Mian et al. [20][a]			93.5
Al-Osaimi et al. [67]			96.5
Li et al. [16]			96.3
Smeets et al. [52]			89.6
Elaiwat et al. [64][a]			97.1
Berretti et al. [26]			≈ 94
Huang et al. [58]			97.6
Wang et al. [5]			**98.4**
Queirolo et al. [27]			**98.4**
Alyuz et al. [50]			97.5
Kakadiaris et al. [3]			97
Faltemier et al. [21]			98.1
Ours			96.3

[a] Only the results on the 3D modal are shown for the corresponding approaches.

Compared to the "neutral vs. nonneutral" experiment, the performance drops are only 3% and 2.2% for rank-1 IR and 0.1% FAR VR, respectively. These results clearly demonstrate that our approach is able to handle large facial expressions caused by an open mouth.

5.3.4. Performance evaluation for robustness against large facial expressions on the SHREC 2008, BU-3DFE and FRGC v2.0 datasets

In this section, the proposed approach has been extensively evaluated on several dataset to test the robustness against large facial expressions. The results of our approach are better than or comparable to the state-of-the-arts in terms of recognition accuracy on three prominent databases (i.e., SHREC 2008, BU-3DFE and FRGC v2.0). The following conclusions can be drawn from the experiential results.

First, a facial scan can be divided into different regions according to their sensitiveness to expressions. Both psychological findings and the 3D FR literature suggest that the mouth area is more sensitive while the other regions are relatively less sensitive to facial expressions. The proposed approach represents a facial scan with a set of local descriptors, which guarantees that a large number of descriptors remain unchanged under facial expressions.

Second, since the proposed local geometrical facial descriptor only considers a small region of the facial scan for feature encoding, it is more sensitive to large facial expressions compared to holistic descriptors. Nevertheless, the proposed approach still achieves competitive performance on the SHREC 2008, BU-3DFE and FRGC v.2.0 datasets when dealing with large facial expressions. Holistic feature descriptors achieve poor performance when dealing with occlusions. Considering the merits of the proposed approach can handle the 3D FR with partial facial data, its performance loss under large facial expressions can be negligible.

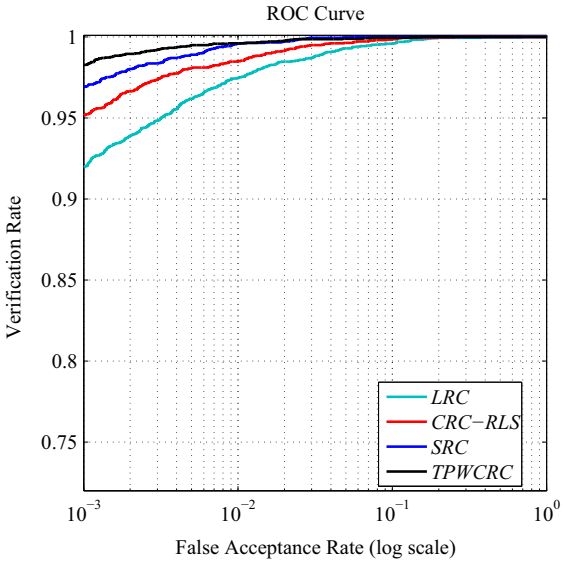

Fig. 17. Comparison to other classification algorithms.

Third, the proposed approach is highly efficient and it enables fast 3D FR since none one-to-one matching between a probe and the gallery faces is required. The computational complexity of the one-to-one matching based approaches. The computational complexity of our approach is further evaluated in Section 5.5.

Lastly, the proposed method is suitable for scenarios with a large gallery set (which usually makes FR a more challenging task). The proposed KMTS–TPWCRC framework incorporates a robust prior knowledge. It not only addresses the single sample problem, but also achieves good performance on large gallery sets.

5.4. Ablative analysis and evaluation of the proposed classification framework

5.4.1. Comparison with other classifiers

In this section, we compare the performance of our classification framework with three competing classifiers including SRC [36], LRC [70,71] and CRC-RLS [13]. It should be noted that our proposed TPWCRC framework can be considered as a variant of CRC-RLS. The major difference between TPWCRC and CRC-RLS is that our TPWCRC framework introduces a locality constraint to deal with the single sample FR problem. Previous work showed that l_2-norm based models can achieve competitive classification accuracy with hundreds of times increase in speed when compared with SRC (l_1-norm based model). The "neutral vs. all" face verification experiment on the FRGC v2.0 dataset (as in Section 5.3.3) is used to perform this comparison, and the results are presented in Fig. 17.

From these results, we can draw the following conclusions. First, the proposed approach generally outperforms SRC, LRC and CRC-RLS in the single sample based 3D FR experiment. Second, the incorporation of the prior classification knowledge as the locality information proves to be useful in dealing with single sample FR problem. Third, SRC, LRC and CRC-RLS achieve a comparable performance, while CRC-RLS and SRC are slightly better than LRC, which are consistent with the results reported in [13,14]. Last, SRC achieves similar classification performance with much more computational complexity. Experimental results show that the time required to classify a single probe scan against 466 individuals in the gallery is 96.57 s, 1.83 s, 1.86 s and 2.41 s for SRC, LRC, CRC-RLS and TPWCRC, respectively.

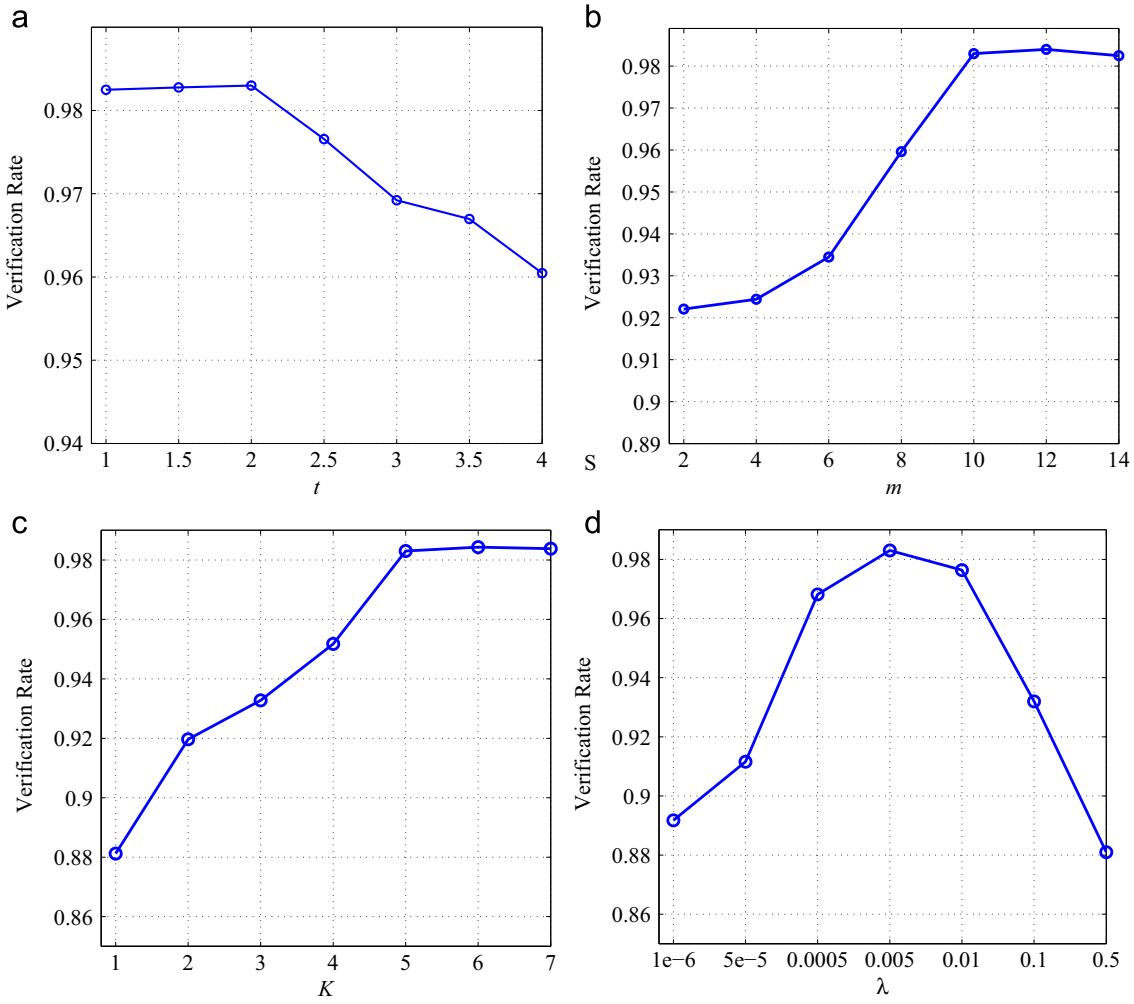

Fig. 18. Performance of the proposed approach for different parameter values.

Table 8
Computational time of the proposed approach for the face identification experiment on the FRGC v2.0 dataset.

Step	Scans	Individuals	Time cost (s)
Face normalization	1	1	0.62
KMTS extraction			5.46
First-phase classification			0.59
	1	466	
Second-phase classification			1.82

5.4.2. Parameter values

All the parameter values (t, m, K and λ) for the proposed approach have been fixed for all of our experiments. In the following part, we present the results of our approach with different parameter values. The "neutral vs. all" face verification experiment on the FRGC v2.0 dataset and the 0.1% FAR VR is used for performance comparison. For each test, the value of one parameter is varied while the other parameters are fixed. The performance of our approach with varying values of t, m, L and λ are shown in Fig. 18. From Fig. 18(a), it can be observed that a good range of values for the parameter t is 1–2. However, a lower value of t will result in an increase in the number of keypoints detected from a facial scan. Consequently, both the memory storage and computational costs are increased. A larger value is therefore preferred for the parameter t. Fig. 18(b) shows that the proposed approach achieves stable performance when the parameter m (the length of each of the four KMTS descriptors) is larger than 10. A large value of m will increase the length of facial descriptors, and thus a smaller value of m is preferred. For the nearest neighbor searching number K, the proposed framework performs well when the value of K is larger than 5, as shown in Fig. 18(c). However, a smaller value of K can result in faster computation. Finally, from Fig. 18 (d) we can see that the performance of the proposed approach is stable when λ is within the range between 0.0005 and 0.01. Given the aforementioned observations, we fixed $t=2$, $m=10$, $K=5$ and $\lambda=0.005$. These parameter values provide a tradeoff between the recognition accuracy and the computational cost.

5.5. Computational performance

Computational cost is an important concern for an efficient and practical 3D FR system, especially for real-world applications. Although previous approaches can perform verification in nearly real-time, the computational cost for face identification is still very high. This becomes even more challenging when the size of the gallery is quite large. That is because, for face identification, a probe scan needs to be matched against all scans in the gallery in a one-to-one manner. In order to analyze the computational complexity of the proposed approach, the timing results of each component of the proposed approach are presented in Table 8. These results were obtained for the "neutral vs. all" face identification experiment on the FRGC v2.0 dataset. The timing experiments were conducted on a PC (with an Intel Core 2 Quad CPU and

16 GB RAM) by MATLAB and C++ hybrid programming. In the facial data preprocessing step, the time required for face normalization and KMTS descriptor extraction is very short (0.62 s and 5.46 s, respectively). During the first-phase classification, the proposed approach uses 0.59 s to calculate the class-based probability estimation for each probe scan. During the second-phase classification, the proposed approach takes 1.82 s to identify each probe scan when the gallery size is 466. We further analyze the computational efficiency of the proposed approach and provide a comparison with the state-of-the-arts in the following paragraph.

Kakadiaris et al. [3] proposed a relatively computationally efficient 3D FR approach among the existing approaches. However, it still takes 15 s for facial data preprocessing. It can perform 1,000 one-to-one face matching per second. As a result, the approach takes 744.2 s to identify 1597 probe scans for the same face identification experiment. Mian et al. [48] proposed an ICP based approach for 3D FR, which requires 4 s to identify a single probe scan with a gallery of 466 faces. Therefore, their approach requires 6,388 s when tested on the same experiment, which is much slower than our proposed approach. Faltermier et al. [21] also proposed an ICP-based 3D FR approach. In their work, 7.5 s are required for data preprocessing and 2.3 s for a single matching. Thus, 1,711,664.6 s are required for the same experimental test, which is very slow for face identification. Wang et al. [5] proposed a computationally efficient 3D FR approach, which requires 0.66 s for matching a probe scan against a gallery of 466 individuals. That approach therefore takes 1054 s for the same face identification experiment, which is the only approach that is faster than our approach for face identification. However, it should be noted that the approach in [5] only works for good quality facial scans. It has not been evaluated for robustness against missing data and few training samples. Ballihi et al. [72] proposed a facial curve analysis based 3D FR approach, which requires 0.68 s to perform a single matching. Therefore, their approach needs 506,057.4 s to perform the above face identification experiment. Drira et al. [15] proposed a 3D PFR approach to address facial expressions, occlusions and pose variations. Their approach required 6.18 s for data preprocessing and 7.45 s for a single matching. Their approach therefore requires 12,293,289.7 s for the aforementioned experiment, which indicates that their approach is computationally infeasible for face identification. Huang et al. [58] proposed a facial surface description using a set of facial depth maps extracted by multi-scale extended Local Binary Patterns (eLBP). Their proposed approach can perform a single matching between a gallery and a probe in around 0.32 s, which required around 238,144 s to perform the aforementioned experiment, making it much slower than the proposed approach. Li et al. [16] proposed a 3D PFR approach based on a curvature-based 3D keypoint detector and three keypoint descriptors, which achieved a highest recognition results on the Bosphorus dataset. However, the performance of their approach highly relies on sophisticated mathematical operations and it identified a single probe using around 75 s (105 scans in the gallery). Thus, their approach required around 119,775 s on the aforementioned experiment (the computational cost should be higher than the reported timing value since there are 466 scans in the gallery for this experiment). This indicates that their approach is subject to improvement in terms of computational efficiency for real-world applications.

These analysis clearly demonstrates the superiority of the proposed approach in terms of computational efficiency when compared to the state-of-the-arts. The proposed approach can perform face identification faster than most existing approaches (all of the 3D PFR approaches). Although it is slightly slower than a few approaches (e.g., [5] and our previous work [12]), it can effectively handle low-quality facial data with missing parts, occlusions and data corruptions. It should be noted that the existing efficient approaches can only work on high-quality facial data. Furthermore, since the computation of each step of the proposed approach is independent, parallel computing techniques can be exploited to further enhance its computational efficiency.

6. Conclusions

In this paper, we proposed an automatic 3D PFR approach using single sample input. The proposed approach represents a 3D face with a set of local geometrical descriptors called KMTS. A TPWCRC framework is proposed to address the single training sample based 3D PFR problem. During the first phase, a class-based probability estimation is computed for each probe scan as a prior classification knowledge. Then, the resulting class-based probability estimation is incorporated into a weighted collaborative representation classification to enhance the locality information for the improvement of its discriminative power. Extensive experiments have been conducted on six challenging 3D facial datasets including the Bosphorus, GavabDB, UMB-DB, SHREC 2008, BU-3DFE and FRGC v2.0 datasets. A promising performance has been achieved on these datasets. Based on these experiments, several conclusions can be drawn. First, keypoint-based local geometrical descriptors can be used to represent a partial 3D face effectively. The proposed descriptor enables robust 3D PFR under missing data, occlusions and data corruptions. Second, a strong prior classification knowledge based on the proposed local descriptor is important for solving the single sample problem. The weighted collaborative representation classification framework incorporated with prior classification knowledge can work in the scenarios of real-world applications. Third, the proposed KMTS-TPWCRC framework is general and can be used for other 3D image classification tasks, e.g., 3D object categorization/classification. The efficiency of the proposed method has been demonstrated through a set of extensive experiments. The experimental results show that the proposed method can effectively handle a number of challenges including occlusions, missing data, expression deformations and limited availability of training samples. Further, the proposed method is highly efficient as it incorporates only simple geometric descriptors and does not involve any one-to-one matching.

Conflict of interest

None declared.

Acknowledgment

This work is supported by the Natural Science Foundation of China under Grant nos. 61403265 and 61471371. This work is also supported by the Science and Technology Plan of Sichuan Province under Grant no. 2015SZ0226.

References

[1] A. Mian, M. Bennamoun, R. Owens, Three-dimensional model-based object recognition and segmentation in cluttered scenes, IEEE Trans. Pattern Anal. Mach. Intell. 10 (2006) 1584–1601.
[2] 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.
[3] I. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, T. Theoharis, Three-dimensional face recognition in the presence of facial expression: an annotated deformable model approach, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 640–649.

[4] Y. Guo, F. Sohel, M. Bennamoun, J. Wan, M. Lu, A novel local surface feature for 3D object recognition under clutter and occlusion, Inf. Sci. 293 (2014) 196–213.
[5] 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.
[6] 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) (2013) 1377–1390.
[7] H. Mohammadzade, D. Hatzinakos, Iterative closest normal point for 3D face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2) (2012) 381–397.
[8] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, Bosphorus database for 3D face analysis, in: Lecture Notes in Computer Science, 2008, 5372:47-56.
[9] A. Moreno, A. Sanchez, GavabDB: a 3D face database, in: Proceedings of Workshop on Biometrics on the Internet, 2004, pp. 75–80.
[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] Y. Lei, M. Bennamoun, A. El-Sallam, An efficient 3D face recognition approach based on the fusion of novel local low-level features, Pattern Recognit. 46 (1) (2013) 24–37.
[12] Y. Lei, M. Bennamoun, M. Hayat, Y. Guo, An efficient 3D face recognition approach using local geometrical signatures, Pattern Recognit. 46 (2) (2014) 509–524.
[13] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: Which helps face recognition?, in: 2011 IEEE International Conference on Computer Vision, ICCV, IEEE, Barcelona, 2011, pp. 471–478.
[14] X. Peng, L. Zhang, Y. Zhang, K. Tan, Learning locality-constrained collaborative representation for robust face recognition, Pattern Recognit. 47 (9) (2014) 2794–2806.
[15] 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.
[16] H. Li, D. Huang, J. Morvan, Y. Wang, L. Chen, Towards 3D face recognition in the real: a registration-free approach using fine-grained matching of 3D keypoint descriptors, Int. J. Comput. Vis. 113 (2) (2014) 128–142.
[17] N. Alyuz, B. Gokberk, L. Akarun, 3-D face recognition under occlusion using masked projection, IEEE Trans. Inf. Forensics Secur. 8 (5) (2013) 789–802.
[18] K. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition, Comput. Vis. Image Underst. 101 (1) (2006) 1–15.
[19] W. Zhao, R. Chellappa, P. Philips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (4) (2003) 399–458.
[20] A. Mian, M. Bennamoun, R. Owens, Keypoint detection and local feature matching for textured 3D face recognition, Int. J. Comput. Vis. 79 (1) (2007) 1–12.
[21] C. Faltemier, K. Bowyer, P. Flynn, A region ensemble for 3D face recognition, IEEE Trans. Inf. Forensics Secur. 13 (1) (2008) 62–73.
[22] K. Chang, K. Bowyer, P. Flynn, Multiple nose region matching for 3D face recognition under varying facial expression, IEEE Trans. Pattern Anal. Mach. Intell. 28 (10) (2006) 1695–1700.
[23] S. Gupta, M. Markey, A. Bovik, Anthropometric 3D face recognition, Int. J. Comput. Vis. 90 (3) (2010) 331–349.
[24] X. Li, T. Jia, H. Zhang, Expression-insensitive 3D face recognition using sparse representation, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Miami, 2009, pp. 2575–2582.
[25] L. Ballihi, B. Ben Amor, M. Daoudi, A. Srivastava, D. Aboutajdine, Selecting 3D curves on the nasal surface using AdaBoost for person authentication, in: Eurographics Workshop on 3D Object Retrieval, 2011, pp. 1–5.
[26] S. Berretti, A. Del Bimbo, P. Pala, 3D face recognition using isogeodesic stripes, IEEE Trans. Pattern Anal. Mach. Intell. 32 (12) (2010) 2162–2177.
[27] C. Queirolo, L. Silva, O. Bellon, 3D face recognition using simulated annealing and the surface interpenetration measure, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2) (2010) 206–219.
[28] N. Alyuz, B. Gokberk, L. Akarun, A 3D face recognition system for expression and occlusion invariance, in: 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, 2008, BTAS 2008, IEEE, Washington DC, 2008, pp. 1–7.
[29] 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.
[30] A. Colombo, C. Cusano, R. Schettini, Detection and restoration of occlusions for 3D face recognition, in: 2006 IEEE International Conference on Multimedia and Expo, IEEE, Toronto, 2006, pp. 1541–1544.
[31] A. Colombo, C. Cusano, R. Schettini, Gappy PCA classification for occlusion tolerant 3D face detection, J. Math. Imaging Vis. 35 (3) (2009) 193–207.
[32] B. Li, A. M. Mian, W. Liu, A. Krishna, Using kinect for face recognition under varying poses, expressions, illumination and disguise, in: 2013 IEEE Workshop on Applications of Computer Vision, WACV, IEEE, Clearwater Beach, 2013, pp. 186–192.
[33] G. Passalis, P. Perakis, T. Theoharis, I. 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.
[34] X. Zhao, E. Dellandrea, L. Chen, I. Kakadiaris, Accurate landmarking of three-dimensional facial data in the presence of facial expressions and occlusions using a three-dimensional statistical facial feature model, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 41 (5) (2011) 1417–1428.
[35] D. Chu, S. Shah, I. Kakadiaris, 3D face recognition for partial data using semi-coupled dictionary learning, in: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG, IEEE, Shanghai, 2013, pp. 1–8.
[36] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[37] C. Chan, J. Kittler, Sparse representation of (multiscale) histograms for face recognition robust to registration and illumination problems, in: 2010 17th IEEE International Conference on Image Processing, ICIP, IEEE, Hong Kong, 2010, pp. 2441–2444.
[38] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, Y. Ma, Toward a practical face recognition system: robust alignment and illumination by sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2) (2012) 372–386.
[39] S. Liao, A. Jain, S. Li, Partial face recognition: alignment-free approach, IEEE Trans. Pattern Anal. Mach. Intell. 35 (5) (2013) 1193–1205.
[40] L. Zhang, Z. Ding, H. Li, J. Lu, 3D face recognition based on multiple keypoint descriptors and sparse representation, Plos One 9 (6) (2014) 1–3.
[41] W. Deng, J. Hu, J. Guo, Extended SRC: undersampled face recognition via intraclass variant dictionary, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1864–1870.
[42] S. Gao, I. Tsang, L. Chia, Kernel sparse representation for image classification and face recognition, in: 2010 European Conference on Computer Vision, Springer, Crete, 2010, pp. 1–14.
[43] Y. Xu, D. Zhang, J. Yang, J. Yang, A two-phase test sample sparse representation method for use with face recognition, IEEE Trans. Circuits Syst. Video Technol. 21 (9) (2011) 1255–1262.
[44] M. Yang, L. Zhang, Gabor feature based sparse representation for face recognition with gabor occlusion dictionary, in: 2010 European Conference on Computer Vision, Springer, Crete, 2010, pp. 448–461.
[45] F. Tombari, S. Salti, L. Di Stefano, Performance evaluation of 3D keypoint detectors, Int. J. Comput. Vis. 102 (1–3) (2013) 198–220.
[46] 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) (2014) 2270–2287.
[47] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: 2010 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, San Francisco, 2010, pp. 3360–3367.
[48] A. Mian, M. Bennamoun, R. Owens, An efficient multimodal 2D and 3D hybrid approach to automatic face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (11) (2007) 1927–1943.
[49] X. Peng, M. Bennamoun, A. Mian, A training-free nose tip detection method from face range images, Pattern Recognit. 44 (3) (2011) 544–558.
[50] 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.
[51] O. Ocegueda, G. Passalis, T. Theoharis, S. Shah, UR3D-C: linear dimensionality reduction for efficient 3D face recognition, in: 2011 IEEE International Joint Conference on Biometrics, IJCB, IEEE, Washington DC, 2011, pp. 1–6.
[52] 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.
[53] S. Berretti, N. Werghi, A. Del Bimbo, P. Pala, Matching 3D face scans using interest points and local histogram descriptors, Comput. Graph. 37 (5) (2013) 509–525.
[54] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, A Riemannian analysis of 3D nose shapes for partial human biometrics, in: IEEE International Conference on Computer Vision, 2009, pp. 2050–2057.
[55] A. Moreno, J. Sanchez, A. Velez, J. Diaz, Face recognition using 3D local geometrical features: PCA vs. SVM, in: Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, IEEE, Zagreb, 2005, pp. 185–190.
[56] M. Mahoor, M. Abdel-Mottaleb, Face recognition based on 3D ridge images obtained from range data, Pattern Recognit. 42 (3) (2009) 445–451.
[57] M. Mousavi, K. Faez, A. Asghari, Three dimensional face recognition using SVM classifier, in: 7th IEEE/ACIS International Conference on Computer and Information Science, IEEE, Portland, 2008, pp. 208–213.
[58] 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.
[59] B. Amberg, R. Knothe, T. Vetter, Expression invariant 3D face recognition with a morphable model, in: IEEE International Conference on Automatic Face and Gesture Recognition, IEEE, Amsterdam, 2008, pp. 1–6.
[60] F. ter Haar, R. Veltkamp, A 3D face matching framework, in: IEEE International Conference on Shape Modelling and Applications, IEEE, Stony Brook, 2008, pp. 103–110.
[61] P. Nair, A. Cavallaro, Registration and retrieval of 3D faces using a point distribution model, in: IEEE International Conference on Shape Modelling and Applications, IEEE, Stony Brook, 2008, pp. 257–258.
[62] D. Xu, P. Hu, W. Cao, H. Li, SHRE'08 entry: 3D face recognition using moment invariants, in: IEEE International Conference on Shape Modeling and Applications, IEEE, Stony Brook, 2008, pp. 261–262.
[63] O. Ocegueda, T. Fang, S. Shah, I. Kakadiaris, 3D face discriminant analysis using Gauss–Markov posterior marginals, IEEE Trans. Pattern Anal. Mach. Intell. 35 (3) (2013) 728–739.
[64] 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.

[65] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, IEEE, San Diego, 2005, pp. 947–954.
[66] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, G. Medioni, Performance of geometrix ActiveID 3D face recognition engine on the FRGC data, in: IEEE Conference on Computer Vision and Pattern Recognition, 2005, p. 154.
[67] 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.
[68] G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici, N. Murtuza, Evaluation of 3D face recognition in the presence of facial expressions: an annotated deformable model approach, in: IEEE Workshop on Face Recognition Grand Challenge Experiments, vol. 3, 2005, pp. 171–179.
[69] J. Cook, C. Mccool, V. Chandran, S. Sridharan, Combined 2D/3D face recognition using Log-Gabor templates, in: IEEE Conference on Video and Signal Based Surveillance, 2006, pp. 83–96.
[70] I. Naseem, R. Togneri, M. Bennamoun, Linear regression for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (11) (2010) 2106–2112.
[71] Q. Shi, A. Eriksson, A. Van Den Hengel, C. Shen, Is face recognition really a compressive sensing problem?, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, Colorado Springs, 2011, pp. 553–560.
[72] L. Ballihi, B. Ben Amor, M. Daoudi, A. Srivastava, D. Aboutajdine, Boosting 3D-geometric features for efficient face recognition and gender classification, IEEE Trans. Inf. Forensics Secur. 7 (6) (2012) 1766–1779.

Yinjie Lei received his M.S. degree from Sichuan University (SCU), China, in the area of image processing, and the Ph.D. degree in Computer Vision from University of Western Australia (UWA), Australia. He is currently an assistant professor at Sichuan University, Chengdu, China. His research interests include image and text understanding, 3D face processing and recognition, 3D modeling, machine learning and statistical pattern recognition.

Yulan Guo received his B.E. and Ph.D. degrees from National University of Defense Technology (NUDT) in 2008 and 2015, respectively. He is currently an assistant professor at NUDT. He was a visiting (joint) PhD student at the University of Western Australia from November 2011 to November 2014. He authored more than 20 peer reviewed journal and conference publications (including IEEE TPAMI, IJCV, PR and IEEE TMM) and one book chapter. He served as a reviewer for more than 10 international journals and several conferences. His research interests include 3D feature extraction, 3D modeling, 3D object recognition, and 3D face recognition.

Munawar Hayat received his Bachelor of Engineering degree from National University of Science and Technology (NUST) in 2009. Later, he was awarded Erasmus Mundus Scholarship for a joint European Master's degree program. He completed his PhD in 2015 from The University of Western Australia (UWA) sponsored by the Scholarship for International Research Fees (SIRF). He is currently a postdoctoral researcher at IBM Research Lab in Melbourne, Australia. His research interests include computer vision, signal and image processing, pattern recognition and machine learning.

Mohammed Bennamoun received the M.S. degree in Control Theory from Queen's University, Kingston, Canada, and the Ph.D. degree in Computer Vision from Queen's/QUT, Brisbane, Australia. He is currently a winthrop professor at the University of Western Australia, Crawley, Australia. His research interests include control theory, robotics, obstacle avoidance, face/object recognition, artificial neural networks, signal/image processing, and computer vision. He published more than 150 journal and conference publications.

Xinzhi Zhou received the B.S. and M.S. degrees from Chongqing University, Chongqing, China, and the Ph.D. degree from Sichuan University, Sichuan, China, in 1988, 1991, and 2003, respectively. He is a professor with the School of Electronics and Information Engineering, Sichuan University. His current research interests include new sensing technology, intelligent system, and intelligent information processing.