An efficient 3D partial face recognition approach with single sample

Yinjie Lei, Siyu Feng, Xinzhi Zhou
College of Electronics and Information Engineering
Sichuan University
Chengdu, Sichuan, 610065, China
Email: yinjie@scu.edu.cn, fsydd@qq.com, xinzhi.zhou@scu.edu.cn

Yulan Guo
College of Electronic Science and Engineering
National University of Defense Technology
Changsha, Hunan, 410073, China
Email: yulan.guo@nudt.edu.cn

***Abstract*—3D partial face recognition under missing parts, occlusions and data corruptions is a major challenge for the practical application of the techniques of 3D face recognition. Moreover, one individual can only provide one sample for training in most practical scenarios, and thus the face recognition with single sample problem is another highly challenging task. We propose an efficient framework for 3D partial face recognition with single sample addressing both of the two problems. First, we represent a facial scan with a set of keypoint based local geometrical descriptors, which gains sufficient robustness to partial facial data along with expression/pose variations. Then, a two-step modified collaborative representation classification scheme is proposed to address the single sample recognition problem. A class-based probability estimation is given during the first classification step, and the obtained result is then incorporated into the modified collaborative representation classification as a locality constraint to improve its classification performance. Extensive experiments on the Bosphorus and FRGC v2.0 datasets demonstrate the efficiency of the proposed approach when addressing the problem of 3D partial face recognition with single sample.**

***Index Terms*—3D partial face recognition, 3D facial representation, locality constraint, collaborative representation, single sample problem**

I. Introduction

Face recognition (FR) has drawn sufficient attention during the last few decades in the community of computer vision and biometric. Compared with other biometric techniques, FR is more suitable for real-world applications due to its non-intrusive and friendly acquisition nature (e.g., access control and video surveillance). According to facial data acquisition modality, FR can be mainly categorized into two folds (i.e., 2D and 3D). Although considerable progress has been made with 2D FR, its performance is still highly challenged by illumination and pose changes. In order to overcome the inherent drawbacks of 2D FR, more researchers have turned to 3D FR which is believed to be more invariant to pose and illumination changes [1], [2], [3], [4], [5]. Moreover, the rapid development of 3D scanning technologies has further boosted the research of 3D FR. In the early years of 3D FR research, most of the 3D facial datasets used for evaluation are acquired in a highly controlled environment, and the resulting facial scans are mostly good in quality and present frontal pose. Most of the existing 3D FR approaches can achieve promising recognition performance based on the above mentioned datasets (e.g., the FRGC v2.0 dataset [6]). Nevertheless, the facial scans can only be acquired with less user cooperation in most practical scenarios, resulting in a partial facial scan (i.e., containing missing parts, occlusions and data corruptions as shown in Fig. 1). Although positive progress has been made by 3D FR, 3D FR from partial facial data (3D PFR) still needs further study to make 3D FR applicable in real-world applications. Recently, the public release of a set of 3D facial datasets which contain the facial scans acquired in uncontrolled environments (e.g., the Bosphorus dataset [7]) has boosted the research on 3D PFR and results in several 3D PFR systems. Moreover, in many real-world applications, an individual can only provide one training sample for FR, which is called the single sample based FR problem. However, most of the existing 2D/3D FR approaches need sufficient training samples per individual to adapt to facial variations (e.g., facial expressions) for the consideration of an accurate FR system. Therefore, the single sample based problem is another major challenge for practical 3D FR applications.

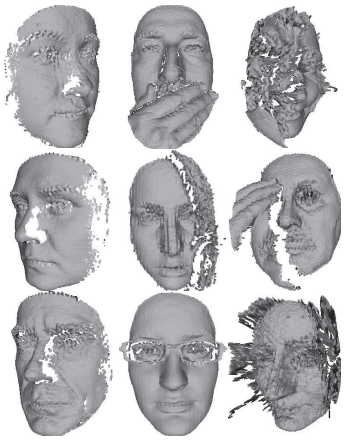


Figure 1: Major challenges for 3D PFR: missing parts (first column), occlusions (second column) and data corruptions (third column).

Aiming to solve the above two major problems of 3D PFR, we propose an efficient 3D FR approach to partial facial data and single sample problem. We first represent a facial

978-1-4673-8644-9/16/$31.00

scan by a set of keypoint-based multiple statistics descriptors, which shows sufficient discriminative power and robustness against partial data, facial expressions and pose variations. On the other hand, a modified collaborative representation classification (CRC) based FR framework is proposed to solve the single sample based FR problem. Specifically, a class-based probability estimation algorithm is proposed to provide a prior classification knowledge to be integrated into the CRC framework as a locality constrained. Compared with the traditional CRC model, our proposed framework can significantly improve the FR performance due to the incorporation of the obtained prior classification knowledge. Our proposed 3D PFR approach is fully automatic and no manual effort is required, and its efficiency to facial variations has been proved by a set of experiments based on FRGC v2.0 and Bosphorus datasets in the experimental section.

II. Related works

In the following, we briefly review the existing approaches most relevant to the proposed approach. Specifically, we will restrain the corresponding review to those approaches to 3D PFR approaches.

In [8], Berretti er al. proposed a meshDOG keypoint detection algorithm along with a local geometric histogram descriptor to represent a 3D face. Then, the most effective features are selected from the obtained local descriptors for 3D PFR. In the similar line, Drira er al. [9] represented a 3D face by a set of radial curves staring from the nosetip. Then, a curve restoration algorithm is introduced to recover the incomplete curve caused by partial facial data. Last, a Riemannian framework was introduced to analyze these curves for an efficient 3D PFR handling expressions, pose variations, missing parts and occlusions. Alyuz et al. [10] proposed the Average Region Models (ARMs) for 3D PFR. First, a facial scan was segmented into several meaningful subregions, followed by a face registration process by separate dense alignments using the obtained ARMs. However, on limitation of their approach is the manual efforts is needed during the segmentation process. Alyuz et al. [11] proposed an automatic 3D PFR system by detecting and remove external occlusions based on an adaptively selected model based registration scheme. Then, the subspace analysis techniques and a masking strategy were applied during the classification phase handling the partial facial data. Colombo et al. [12] first calculated a mean face, and the facial external occlusions were detected using the difference between an input face and the corresponding mean face. Then, the detected occlusions were removed and recovered and followed by a Gappy PCA based scheme to discriminate a face from non-face images even when only partial facial data available. Similarly, Colombo et al. [13] detected facial external occlusions based on the difference between an original face and its corresponding eigenface approximation. The regions which were decided as occlusions were recovered by using the Gappy PCA algorithm. Last, the FR was performed using a modified Fisherfaces algorithm based framework. Passalis et al. [14] introduced an Annotated Face Model (AFM) for facial scan registration and fitting. Then, the facial symmetry was used to handle the partial facial data problem during the registration and fitting process, and the wavelet-based biometric signatures were applied to perform 3D PFR during the classification process. Li et al. [15] proposed a preprocessing algorithm to transform a facial scan to a canonical frontal view independent to their initial pose even for profile faces. Then, the corresponding canonical face was codded by a sparse representation algorithm, and the input face was recognized as the identity with the smallest reconstruction error using sparse coding. Li et al. [16] detected facial keypoints using curvature-based detector, and three sophisticated feature descriptors were extracted from the detected keypoints. Then, a multi-task sparse representation based fine-grained matching scheme was proposed to perform 3D PFR. They provided the best recognition performance on the Bosphorus dataset. Zhao et al. [17] proposed a facial landmark localization algorithm for 3D faces using a statistical facial feature model. The proposed algorithm is robust against facial partial data and can effectively address 3D PFR problem.

III. Multiple keypoint based 3D facial descriptor

A. Keypoint detection from a facial scan

The aim of the keypoint detection algorithm is to determine a set of highly repeatable and discriminative salient points from a facial surface. In this paper, we apply a 3D facial keypoint detection approach first proposed in [18] by Mian et al, which gains superior detection performance [19], [20]. Specifically, these 3D keypoints can be repetitively detected on 3D face under different variations (e.g., noises, expressions, poses and occlusions). Moreover, the applied 3D keypoint approach only involves simple calculation without any complicated mathematical operations. We here briefly describe the keypoint detection approach in below.

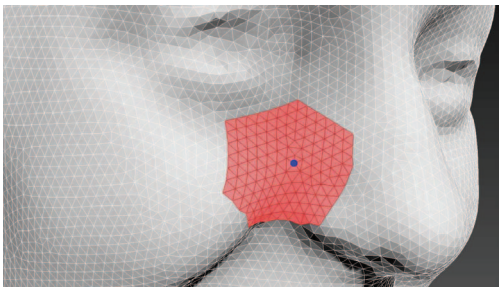


Figure 2: An illustration of a detected keypoint and its corresponding local patch on a facial surface.

The input of the keypoint detection module is a facial range image $\mathbf{R}$ and its corresponding pointcloud $\mathbf{F} = [x_i, y_i, z_i]^T$ (where $i = 1, \ldots, n$, which indicates the total number of points). We then resample $\mathbf{F}$ with an uniform intervals of 4 mm in both (x, y) directions, resulting in the sample point p to be considered as a keypoint candidate for further computation. Next, we crop a local patch $\mathbf{L} = [x_j, y_j, z_j]^T$ around candidate p (set as the center of the corresponding patch) by a sphere

with a radius of $r = 20\,\mathrm{mm}$. The salience of the local patch $\mathbf{L}$ can be used to determine whether p is a keypoint or not. The mean vector $\mathbf{m}$ of patch $\mathbf{L}$ along with its corresponding covariance matrix $\mathbf{C}$ can be calculated as:

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

where $\mathbf{L}_j$ represents the jth column of $\mathbf{L}$. Then, the diagonal matrix of the eigenvlues $\mathbf{D}$ and the corresponding eigenvector matrix $\mathbf{V}$ of the covariance matrix $\mathbf{C}$ can be obtained by using Principal Component Analysis (PCA) as:

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

The principal axes of $\mathbf{L}$ can be calculated using Hotelling transform as: $\mathbf{L}'_j = \mathbf{V}(\mathbf{L}_j - \mathbf{m})$. Assuming $\mathbf{L}_x$ and $\mathbf{L}_y$ to be the x and y components of the principal axes of local patch $\mathbf{L}$ (where $\mathbf{L}'_x = x_j$ and $\mathbf{L}'_y = y_j$).

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

where ζ indicates the difference between the first two principal axes of $\mathbf{L}$. The value of ζ reflects the shape variation of local patch $\mathbf{L}$, i.e., ζ is zero if $\mathbf{L}$ is planar or spherical, and ζ has a non-zero value if the $\mathbf{L}$ is unsymmetrical. Moreover, its value is proportional to such shape variation. Based on the above observation, if the value of ζ is larger than a threshold (i.e., $\zeta \geq t$), its corresponding point p is considered as a keypoint. In this paper, we empirically choose $t = 2\,\mathrm{mm}$. Figure. 2 shows an example of a detected keypoint and its corresponding local patch on a facial surface.

B. Multiple triangle statistics based local descriptor

We represent the local patch $\mathbf{L}$ around each detected keypoint p by a set of multiple triangle statistics as local feature which is based on our previous work [4]. The local patch $\mathbf{L}$, centered at the detected keypoint, is represented by multiple spatial triangles as shown in Fig. 3 (readers can refer to [4] for details). Specifically, one vertex is the detected keypoint p and the remaining two vertices are randomly picked up from the local patch $\mathbf{L}$ for each of the spatial triangles. Next, we design four types of geometrical features based on those extracted multiple spatial triangles as follows. (1) A: the angle between the two lines determined by the two random vertices and the keypoint. (2) C: the radius of the circle circumscribed to each spatial triangle. (3) D: the length of the line between the two random vertices. (4) N: the angle between the z-axis and the line determined by the two random vertices. An illustration of the designed descriptor is shown in Fig. 3. Specifically, if $\mathbf{L}$ contains n vertices except the keypoint, our local descriptor results in n^2-2n spatial triangles, and thus the dimensionality of the feature vector extracted from the corresponding local patch $\mathbf{L}$ is also $n^2 - 2n$. Then, the obtained feature vector is normalized into the range of $[-1, +1]$ and quantize the each of the four statistics into a histogram description of m bins ($m = 30$ in our case). Finally, we concatenate all the four histogram based descriptors into one feature vector with a length of $m \times 4$.

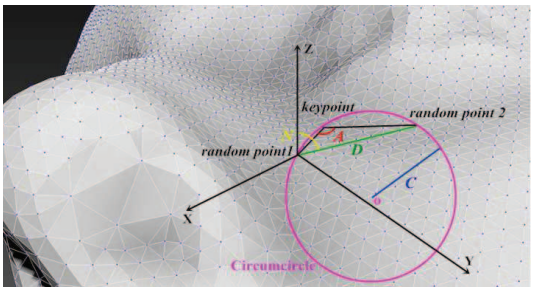


Figure 3: An illustration of the proposed four types of multiple triangle statistics based on one spatial triangle on a facial surface.

IV. A TWO STEP BASED 3D PARTIAL FACE RECOGNITION

A. The first classification step

In this work, we propose a locality preservation algorithm incorporated into a modified CRC model in order to improve its recognition performance. To the best of our knowledge, only few existing approaches have been proposed to investigate the element-based locality preservation strategies to improve the local consistency in the linear coding scheme (e.g., [21]). We therefore propose a class-based locality preservation scheme to adapt to our single sample based 3D PFR. Compared with the traditional element-based locality preservation strategy, the proposed approach can provide a more reliable locality information by considering all the keypoint base local feature descriptors. In contrast, the traditional approaches mainly used holistic descriptors, and thus had only one chance for probability estimation when accommodating he single sample based 3D PFR.

First, a simple 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 keypoint based local descriptor) over the entire dictionary, where $\mathbf{D}_c$ are all the training descriptors for the cth individual.

The resulting probability estimation is a probability vector $\mathbf{p} = [\mathbf{p}(\mathbf{Y}|\mathbf{D}_1), \mathbf{p}(\mathbf{Y}|\mathbf{D}_2), \ldots, \mathbf{p}(\mathbf{Y}|\mathbf{D}_C)]$, with C entries (where C is the number of all individuals in the training set). Each entry of the probability vector indicates the prior possibility for classifying the input 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:

$$\mathbf{d}_c^i = \min(\| (\mathbf{y}_i, \mathbf{D}_c) \|) \tag{4}$$

$$\mathbf{d}_c = \sum_{i=1}^{n}\mathbf{d}_c^i \tag{5}$$

where $\|\cdot\|$ represents the Euclidean distance, $\mathbf{d}_c^i$ denotes the minimal distance between the ith feature descriptor $\mathbf{y}_i$ of the

input test (corresponds to the the ith keypoint) 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 in the training set. Last, a probability vector $\mathbf{p}$ can be computed as:

$$\mathbf{p}(\mathbf{Y}|\mathbf{D}_c) = \exp(\frac{\mathbf{d}_c/\mathbf{d}_{\min}}{\varepsilon}) \tag{6}$$

where $\mathbf{d}_{\min} = \min(\mathbf{d})$, ε is a constant. The proposed probability estimation algorithm enhances the locality information when handling single sample based FR problem. The output of the first step classification can provide a locality constraint for the modified CRC model during the second step classification.

B. The second classification step

In this paper, a facial scan is represented by a set of keypoint based local descriptors (both training and test samples). Therefore, the samples from the training set will result in thousands of atoms for those linear coding based approaches (e.g., CRC and SRC), which is determined by the number of individuals. Moreover, the extracted keypoint based local descriptors lack similarity (compared with holistic descriptor) even for those descriptors from the same individual. Consequently, we mange to filter the training descriptors in order to select the most relevant ones according to the input test. First, we apply the K-Nearest Neighbor (K-NN) search (with Euclidean distance) on the training samples of all the C individuals, resulting in a subdictionary, $\mathbf{D}_c = [\mathbf{D}_c^1, \mathbf{D}_c^2, \ldots, \mathbf{D}_c^K]$ (belong to the cth individual). Note that, K is set to be a constant in this work. As a result, the subset for all the training dictionary can be generated as $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_C]$. Compared to the original training dictionary, the resulting subset is more compact and relevant to the input test.

Some of the existing work argue that the locality is more important than sparsity due to the locality will lead to sparsity but sparsity may not result in locality [21]. Therefore, we incorporate the learned class-based probability estimation $\mathbf{p}$ as a weighing term into the modified CRC framework in order to improve its locality. In detail, given one keypoint based local descriptor $\mathbf{y}_i$, the objective function of our modified CRC framework is as follow:

$$\hat{\mathbf{x}}_i = \arg\min_{\mathbf{x}_i} \parallel \mathbf{D}\mathbf{x}_i - \mathbf{y}_i \parallel_2^2 + \lambda \parallel \mathbf{p}^T \mathbf{x}_i \parallel_2^2 \tag{7}$$

where $\mathbf{x}_i$ is the coding coefficient for the ith local descriptor $\mathbf{y}_i$ of the input test. Our modified CRC framework will select based from local neighbors to ensure locality for the coding results. The incorporation of the probability estimation as weighting term can significantly improve the recognition performance. Moreover, the solution of our modified CRC framework can be derived analytically without any computational demanding optimization algorithms. Finally, the reconstruction residual to the specific individual can be used to determine the recognition of the input test:

$$r_c(\mathbf{Y}) = \frac{1}{n}\sum_{i=1}^{n} \parallel \mathbf{y}_i - \mathbf{D}_c\delta_c(\hat{\mathbf{x}}_i) \parallel_2 / \parallel \delta_c(\hat{\mathbf{x}}_i) \parallel_2 \tag{8}$$

where $\delta_c(\cdot)$ is selection function correlates to the coefficients with the cth individual. Equation 8 uses a sum pooling scheme by taking all the keypoint based local descriptors into consideration with respect to each individual. As a result, the total reconstruction residuals $r = (r_1, r_{2,\ldots}, r_C)$ can be viewed as a similarity score (negative polarity).

For face verification, given a threshold η:

$$\begin{array}{ll} \text{Accept} & \text{if } \eta - r_c(\mathbf{Y}) \geq 0 \\ \text{Reject} & \text{if } \eta - r_c(\mathbf{Y}) < 0 \end{array}. \tag{9}$$

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:

$$I(\mathbf{Y}) = \arg\max_{1\leq c\leq C}(r_c(\mathbf{Y})), \tag{10}$$

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

V. Experimental results

A. Data preprocessing

Most of the raw facial data from both Bosphorus and FRGC v2.0 datasets contain noise spikes, holes, pose variations as well as external occlusions (e.g., by clothes, hair and hand). Obviously, a data preprocessing procedure is required aiming to improve the recognition performance. The data preprocessing includes four steps, which are explained as follow.

First, the nosetip is automatically detected from the facial surface using an efficient algorithm proposed by Mian et al. in [22]. Based on the detected nosetip, we crop the 3D face by discarding the points outside a sphere centered at the nosetip with a radius of 80 mm. Next, the noise spikes are eliminated by removing the outlier vertices, and the missing parts are automatically identified by locating boundary edges, iteratively linking them into loops, and performing triangulation on the obtained loops. Third, the pose correlation algorithm proposed in [22] is applied to correct the pose of a 3D face to its frontal view. Last, the preprocessed facial mesh is resampled onto a square grid of 161×161 mm with a uniform resolution of 1 mm.

B. Experimental results on the Bosphorus dataset

The Bosphorus dataset includes more than 4,000 facial scans present different poses, expressions and external occlusions. Specifically, the occlusions have four types, i.e., (1) mouth occlusion by hand, (2) glasses, (3) face occlusion by hair, and (4) eye/forehead occlusion by hand. With the consideration of a fair comparison, we follow the same evaluation protocol as in the literature [9]. In detail, one neutral facial scan of each individual is selected as training set. Next, all the 381 facial scans with above mentioned four types of external

occlusions are selected as the testing set. The resulting rank-1 Identification Rate (IR) with different experimental setups are reported in Fig. 4, and an overall rank-1 IR of 92.7% is achieved by our approach. Moreover, the comparison between our approach and the state-of-the-arts is shown in Table I.

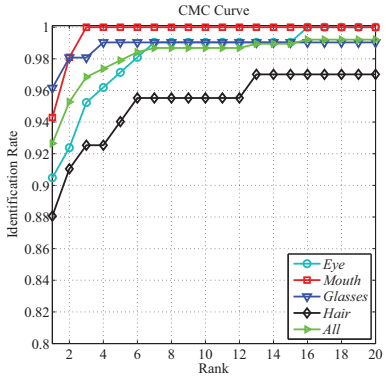


Figure 4: The CMC curves of the face identification experiments with different occlusions on the Bosphorus dataset.

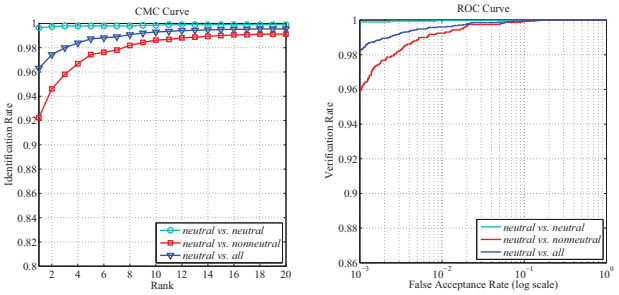
The approach proposed by Li et al. [16] reported the highest overall rank-1 IR of 99.2% on the Bosphorus dataset. They used two complicated keypoint detection algorithms along with three keypoint based local descriptors to represent a facial scan. However, the major limitation of their approach is that it is computationally very expensive, and thus is difficult to be applied in real-world scenarios. In contrast, the proposed approach involves only a set of low-level simple descriptors and thus is computationally efficient. Compared with [16], the proposed approach is almost 10 times faster than their approach during the feature extraction phase, and is 2 times faster during the classification phase. Moreover, it is believed that the recognition performance of the proposed approach can be further improved if more sophisticated features are used to represent a facial scan as [16]. However, we propose a simple 3D PFR framework addressing the single sample problem and can be applied in the real-world scenarios.

Table I: The rank-1 IRs under different occlusions on the Bosphorus dataset.

approaches	eye	mouth	glasses	hair	overall
Alyuz et al. [10]	93.6%	93.6%	97.8%	89.6%	93.6%
Colombo et al. [23]	91.1%	74.7%	94.2%	90.4%	87.6%
Drira et al. [9]	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%

C. Experimental results on the FRGC v2.0 dataset

The FRGC v2.0 dataset is one of the largest publicly available 3D facial dataset, which contains 4,007 3D facial scans belonging to 466 individuals. Specifically, 2,410 facial scans are with nearly neutral expressions and the remaining scans are with different facial expressions. Following the same experimental setups as the literature, we select the first neutral scan of each individual to generate the training set (466 in total). Then, three experimental subsets are generated: (1) "neutral vs. neutral" experiment (1,944 probes); (2) "neutral vs. nonneutral" experiment (1,597 probes); (3) "neutral vs. all" experiment (3,541 probes). The CMC and ROC curves of the proposed approach on the above mentioned experimental subsets are illustrated in Fig. 5. The proposed approach obtains an overall rank-1 IR of 96.3% and a 0.1% FAR VR of 98.3%. A rank-1 IR of 92.2% and a 0.1% FAR VR of 96% are achieve by the proposed approach in the most challenging subset, i.e., "neutral vs. nonneutral". In the simpler case of "neutral vs. neutral", the proposed approach achieves a rank-1 IR of 99.6% and a 0.1% FAR of 99.9%, respectively.



(a) (b)

Figure 5: The CMC and ROC curves for the "neutral vs. neutral", "neutral vs. nonneutral" and "neutral vs. all" experiments on the FRGC v2.0 dataset. (a) the CMC curves and (b) the ROC curves.

The comparison between the proposed approach and the existing approaches with respect to recognition performance (i.e., 0.1% FAR VR) is presented in Table II. The listed approaches for comparison are specifically proposed for the FRGC v2.0 dataset. However, it is not difficult to achieve a high recognition performance by most of the existing 3D FR approaches since the facial scans are in relatively high quality when compared with the Bosphorus dataset, and thus the FRGC v2.0 is not suitable for evaluating the 3D PFR approaches. Although the proposed approach is not specifically developed for the FRGC v2.0 dataset, it still achieves a very competitive performance (a 0.1% FAR VR of 98.3%) in the "neutral vs. all" subset, which is comparable to the highest reported result (a 0.1% FAR VR of 98.6% reported by [24]).

Table II: Comparison with the state-of-the-art on the FRGC v2.0 dataset with respect to 0.1% FAR VRs for the "neutral vs. neutral", "neutral vs. nonneutral" and "neutral vs. all" subsets.

approaches	*n vs. n*	*n vs. nn*	*n vs. a*
Berretti et al. [25]	97.7%	91.4%	95.5%
Al-Osaimi et al. [26]	98.4%	97.8%	98.1%
Wang et al. [24]	-	-	**98.6%**
Huang et al. [27]	99.6%	97.2%	98.4%
Ours	**99.9%**	96%	98.3%

VI. Conclusions

In this work, a fully automatic and efficient 3D PFR approach with single training sample is proposed. First, a facial scan is represented by a set of keypoint based local geometrical descriptors. Then, based on the extracted descriptors a modified two-step CRC framework which enhances the locality constraint to handle the single training sample problem. During the first step classification, a probability estimation for each input test to indicate the prior classification knowledge. Then, the resulting probability estimation is incorporated into a modified CRC framework as a locality constraint during the second classification step. It is observed that, the recognition performance of the proposed framework has been significantly improved due to the incorporation of the probability estimation. The proposed approach has been evaluated based on two challenging 3D facial datasets (i.e., Bosphorus and FRGC v2.0), and a promising performance has been achieved. Several conclusions can be drawn from our experimental results. First, compared with holistic representation, our keypoint based local geometrical descriptor is more suitable to addressing 3D PFR under missing parts, occlusions and data corruptions. Second, the probability estimation during the first classification step can significantly improve the recognition performance when dealing with single sample problem. Third, the proposed classification approach has a strong ability of generalization and can be extended to other classification tasks, e.g., 3D object classification, 3D fingerprint classification. Furthermore, the proposed approach is computationally efficient without any complicated mathematical operations and thus can be applied directed in the real-world scenarios.

Acknowledgment

This work is supported by the Natural Science Foundation of China under Grant Numbers 61403265 and 61471371. This work is also supported by the Science and Technology Plan of Sichuan Province under Grant Number 2015SZ0226.

References

[1] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan, "Rotational projection statistics for 3D local surface description and object recognition," *International Journal of Computer Vision*, vol. 105, no. 1, pp. 63–86, 2013.

[2] Y. Guo, F. Sohel, M. Bennamoun, J. Wan, and M. Lu, "A novel local surface feature for 3d object recognition under clutter and occlusion," *Information Sciences*, vol. 293, pp. 196–213, 2014.

[3] Y. Guo, F. Sohel, M. Bennamoun, J. Wan, and M. Luu, "An accurate and robust range image registration algorithm for 3d object modeling," *IEEE Transactions on Multimedia*, vol. 16, no. 5, pp. 1377–1390, 2013.

[4] Y. Lei, M. Bennamoun, and A. El-Sallam, "An efficient 3D face recognition approach based on the fusion of novel local low-level features," *Pattern Recognition*, vol. 46, no. 1, pp. 24–37, 2013.

[5] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, "An efficient 3D face recognition approach using local geometrical signatures," *Pattern Recognition*, vol. 46, no. 2, pp. 509–524, 2014.

[6] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2005, pp. 947–954.

[7] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3d face analysis," in *Biometrics and Identity Management*. Springer, 2008, pp. 47–56.

[8] S. Berretti, N. Werghi, A. del Bimbo, and P. Pala, "Selecting stable keypoints and local descriptors for person identification using 3D face scans," *The Visual Computer*, vol. 30, no. 11, pp. 1275–1292, 2014.

[9] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D face recognition under expressions, occlusions and pose variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2270–2283, 2013.

[10] N. Alyuz, B. Gokberk, and L. Akarun, "A 3d face recognition system for expression and occlusion invariance," in *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*. IEEE, 2008, pp. 1–7.

[11] N. Alyuz, B. Gokberk, and L. Akaru, "3-d face recognition under occlusion using masked projection," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 5, pp. 789–802, 2013.

[12] A. Colombo, C. Cusano, and R. Schettini, "Gappy pca classification for occlusion tolerant 3D face detection," *Journal of mathematical imaging and vision*, vol. 35, no. 3, pp. 193–207, 2009.

[13] A. Colombo, C. Cusano, and R. Schettin, "Detection and restoration of occlusions for 3D face recognition," in *2006 IEEE International Conference on Multimedia and Expo*. IEEE, 2006, pp. 1541–1544.

[14] G. Passalis, P. Perakis, T. Theoharis, and I. Kakadiaris, "Using facial symmetry to handle pose variations in real-world 3d face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1938–1951, 2011.

[15] B. Li, A. M. Mian, W. Liu, and 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, 2013, pp. 186–192.

[16] H. Li, D. Huang, J. Morvan, Y. Wang, and L. Chen, "Towards 3D face recognition in the real: A registration-free approach using fine-grained matching of 3D keypoint descriptors," *International Journal of Computer Vision*, vol. 113, no. 2, pp. 128–142, 2014.

[17] X. Zhao, E. Dellandrea, L. Chen, and 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 Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 5, pp. 1417–1428, 2011.

[18] A. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3D face recognition," *International Journal on Computer Vision*, vol. 79, no. 1, pp. 1–12, 2007.

[19] F. Tombari, S. Salti, and L. Di Stefano, "Performance evaluation of 3d keypoint detectors," *International Journal of Computer Vision*, vol. 102, no. 1-3, pp. 198–220, 2013.

[20] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan, "3D object recognition in cluttered scenes with local surface features: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2270–2287, 2014.

[21] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 3360–3367.

[22] A. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D and 3D hybrid approach to automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1927–1943, 2007.

[23] A. Colombo, C. Cusano, and R. Schettini, "Three-dimensional occlusion detection and restoration of partially occluded faces," *Journal of mathematical imaging and vision*, vol. 40, no. 1, pp. 105–119, 2011.

[24] Y. Wang, J. Liu, and X. Tang, "Robust 3D face recognition by local shape difference boosting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1858–1870, 2010.

[25] S. Berretti, A. Del Bimbo, and P. Pala, "3D face recognition using isogeodesic stripes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2162–2177, 2010.

[26] F. Al-Osaimi, M. Bennamoun, and A. Mian, "An expression deformation approach to non-rigid 3D face recognition," *International Journal of Computer Vision*, vol. 81, no. 3, pp. 302–316, 2009.

[27] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "3-d face recognition using elbp-based facial description and local feature hybrid matching," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1551–1565, 2012.