Robust 3D Local SIFT Features for 3D Face Recognition

Yue Ming[1(✉)] and Yi Jin[2]

[1] School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, People's Republic of China
myname35875235@126.com
[2] School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, People's Republic of China

Abstract. In this paper, a robust 3D local SIFT feature is proposed for 3D face recognition. For preprocessing the original 3D face data, facial regional segmentation is first employed by fusing curvature characteristics and shape band mechanism. Then, we design a new local descriptor for the extracted regions, called 3D local Scale-Invariant Feature Transform (3D LSIFT). The key point detection based on 3D LSIFT can effectively reflect the geometric characteristic of 3D facial surface by encoding the gray and depth information captured by 3D face data. Then, 3D LSIFT descriptor extends to describe the discrimination on 3D faces. Experimental results based on the common international 3D face databases demonstrate the higher-qualified performance of our proposed algorithm with effectiveness, robustness, and universality.

Keywords: 3D face recognition · 3D Local Scale-Invariant Feature Transform · Facial region segmentation · Depth information

1 Introduction

Biometrics systems have been presented for several decades with wide applications, such as 3D movies, human-machine interaction, and intelligent monitoring [1]. Among them, face recognition has a high-level preference for a huge number of researchers and organizations, mainly because of its non-invasiveness and user-friendliness. However, face recognition developed by 2D images has been hindered by the obstacles induced by pose, illumination, expressions, and other varied characteristics in real-world situations. With the rapid development of 3D digital capturing devices, facial recognition in 3D data has been introduced to solve the challenging issues using a variety of methods.

A sufficient broad investigation of face recognition has been provided in [2], specifically on 3D face recognition. Empirical study shows that facial shape has significant variations in terms of different regions on the facial surface. In order to effectively encode facial anatomical structure and describe the discriminant features, we introduce segmenting scheme to address the facial region attributes

H. Liu et al. (Eds.): ICIRA 2015, Part III, LNAI 9246, pp. 352–359, 2015.
DOI: 10.1007/978-3-319-22873-0_31

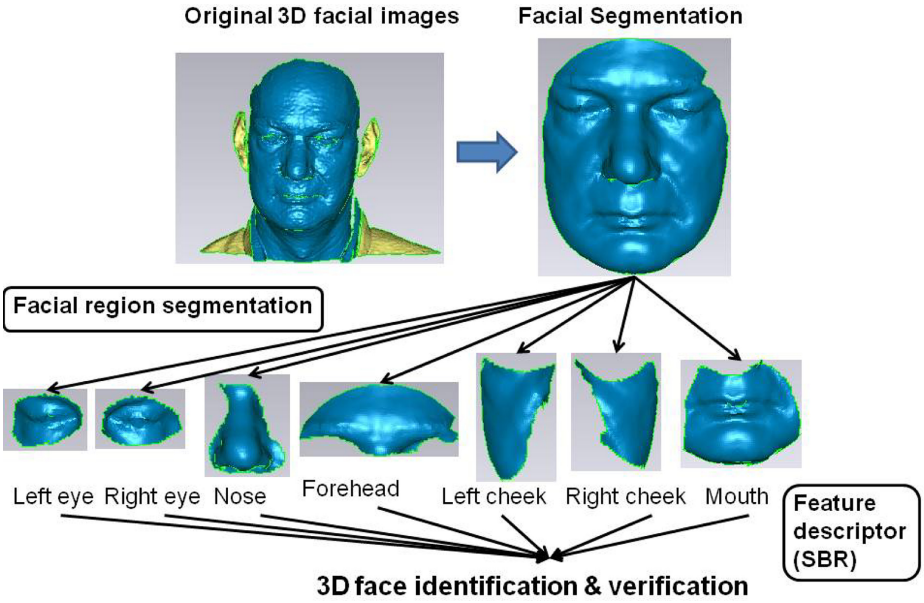

Fig. 1. The framework of our proposed 3D face recognition system

and develop a new recognition framework to 3D facial data. Our framework for 3D face recognition is composed of three parts: facial region segmentation, feature detection and description, and 3D face recognition as shown in Figure 1.

In our framework, a group of facial local regions can be coarsely located by curvature information and shape band [3] algorithm is introduced to refine the localizations which possess more discriminant power. Then, we exploit the original Scale-Invariant Feature Transform [4] to 3D local surface on the sub-regions of 3D face data, and encode the gray and depth data for the detected key point description. 3D local Scale-Invariant Feature Transform (3D LSIFT) can be extracted to describe the regional characteristics for achieving the 3D face recognition. Our method is more robust to image artifacts, lighting variance, winkles, and occlusions and corruptions and shows the generalization based on the challenging databases.

The organization of the paper is as follows. Facial regional segmentation is described in Section 2. Section 3 proposes 3D local Scale-Invariant Feature Transform (3D LSIFT). Section 4 reports the experimental results based on the challenging 3D face databases and compares the performance of algorithms. Finally, the paper is concluded in Section 5.

2 3D Facial Regional Segmentation

The original 3D facial images in the databases usually contain some non-facial areas, such as ears, necks and shoulders as shown in Figure 1. The common

3D face databases mainly provide the 2D texture image and its corresponding valid point matrix with 3D face images simultaneously. By corresponding valid point on 2D and 3D images, the facial area is coarsely detected. Exploiting our previous research [5][6], the facial registration is calculated by Axis-angle representation, which can align the input with the reference model fixed. Then, with the different values and directions of curvature [7], the different areas of a face can be coarsely detected.

In this paper, our refined region extraction exploits shape band detection [3] on 3D facial images in the spherical domain. Given a 3D facial image, we calculate the shape index [8] values based on the curvature as shown in Figure 2 and we choose left/right inner eye corner points, nasal tip point and left/right nasal basis points as five facial key points. The key points can be treated as regional centers for region matching with reference template. Based on facial cropping parameters [9] and regional shape characteristics, we can obtain a series of regional centers $C = [c_1, \cdots, c_n]$ (n is the number of regions) and the corresponding regional radius $r = [r_1, \cdots, r_n]$. We first match a region template $P = \{p_1, \cdots, p_m\}$ ($p_i (i = 1, \cdots, m)$ is the samples points of an average face model T) to a shape index image $SI = \{si_1, \cdots, si_n\}$ expressed as a set of contour fragments. We translate the template to the coordinate system centered at the corresponding regional centers, and construct the shape band $SB(P)$ as follows,

$$SB(P(c)) = \{ip \in T | \exists p_i \| ip - (p_i + c_i)\|_2 \leq r\} \tag{1}$$

where $\|\cdot\|_2$ denotes the l_2 norm.

The detection and matching of the input face model I can be treated as the selection of the optimal central points $C^* = [c_1^*, \cdots, c_n^*] \in I$ and a subset $SI^* = [si_1^*, \cdots, si_n^*]$ as follows,

$$D(P(c^*), SI^*) = \min_{c \in T, SI' \subset SI} D(P(c), SI') \tag{2}$$

where D is a shape band distance defined in literature [3].

Then, Shape Context distance SC is used to the fine process to determine the optimal segmentation. For each shape index segment si_i in SI, its minimum distance can be introduced to select the regions with the preset thresholds Th_i (the value depends on the different databases). If the minimum distance defined in Equation (2) is less than the threshold, then si_i is adjunct to $P(c_i)$, denoted the adjunct segments as $AS(C_i)$,

$$\min Dist(si_i, P(c_i, \varepsilon_i)) = \min_{p_i \in P, q \in Q} \|q - (\varepsilon_i p_i + c_i)\|_2 \tag{3}$$

where Q is the point set in the template segment si_i and ε_i is the scale value of the selected regions. SC can be calculated by the distance between the shape index segments and the template position. We treat the minimal shape distance as the matching segments of the input face model aligned with the regional template.

$$SD(c_i) = \min_{S \subseteq AS(c_i)} SC(US, P(c_i)) \tag{4}$$

The final detected regions of the input face model are the one with the smallest shape distance as shown in Figure 2.

$$region = \arg\min_{c_i \in C} SD(c_i) \tag{5}$$

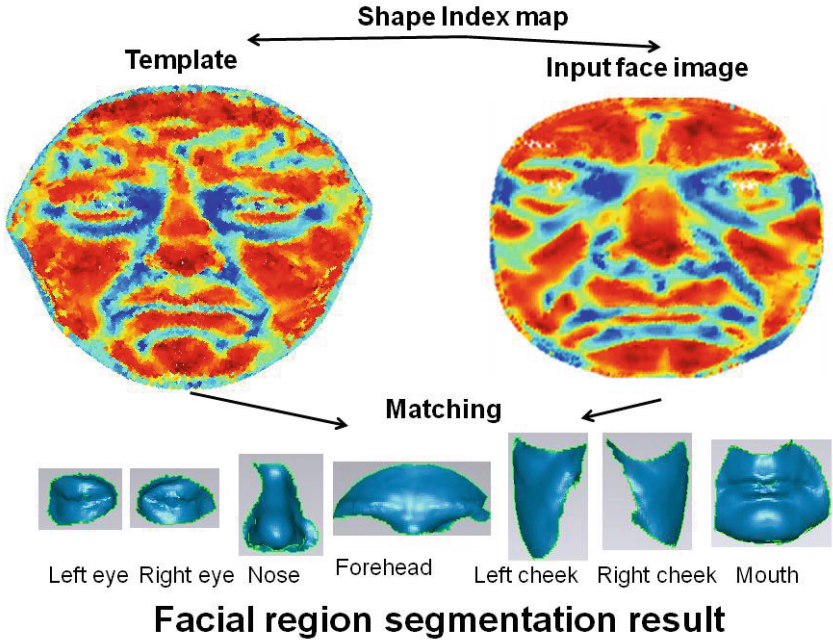

Fig. 2. 3D facial regional segmentation

3 3D Local Scale-Invariant Feature Transform (3D LSIFT)

Inspired by the facial geometric attributes with depth information and scale invariant descriptors for the task of discrimination, a novel descriptor is proposed for a group of facial regions, denoted as 3D Local Scale-Invariant Feature Transform (3D LSIFT). Our feature extraction can be divided into two steps, e.g., the key point detection and description. For the detection, density-invariant Gaussian filters are defined to calculate the filtered mesh sets on the facial surface geometry. We only use the corresponding mesh image of 3D face image and employ with the similar way as SIFT to fix the positions of the key points, which are the local extreme of the Gaussian pyramid.

The second step for 3D LSIFT is to build the descriptor for each key point on the different facial regions. Since the key point detection is the derivative of the original SIFT algorithm, the detected points is with the invariance on scale and rotation. In order to extend the invariance to 3D space, the depth information

around the key points is added to describe such points. Suppose the key point is positioned at (i, j), and the gray image is denoted by I, and the depth image as D. For each pixel with in the local facial region, we compute three gradients for it,

$$\begin{aligned} I_x &= I(i+1, j) - I(i, j) \\ I_y &= I(i, j+1) - I(i, j) \\ D_z^x &= D(i+1, j) - D(i, j) \\ D_z^y &= D(i, j+1) - D(i, j) \end{aligned} \tag{6}$$

In order to describe both the facial gray and depth data into our 3D LSIFT descriptor, the magnitude and orientation on each facial pixel neighboring the key point is calculated in the corresponding scale space of the Gaussian pyramid. For xy, xz, and yz planes, the gradient's magnitude and orientation of each pixel are calculated by I_x and I_y, I_x and D_z^x, I_y and D_z^y, respectively. We divide the local facial region with size 16*16 into 16 grids around the key points in the corresponding scale space. For each grid with 4*4 pixels, the facial rotation can be quantized into eight directions. And then for each orientation, the magnitude is summed up into the corresponding bins of the histogram. We normalize the summation of the histogram to 1. By fusing the interesting point descriptors on the different facial regions, each bin can be quantized into [0, 255]. Thus, our 3D LSIFT feature vector has 4*4*8*3=384 dimensions.

4 Experiments

In the experimental section, we test the performance of our 3D face recognition framework in both identification and verification scenarios. From the preprocessed images, we can extract the different discriminative features to represent the individuals and compare the accuracy with other popular methods based on the same application purpose. The similarity measure is Euclidean distance.

4.1 Comparison Evaluation for Verification Scenarios

In these experiments, we present the comparative evaluations on FRGC v2.0 3D face databases [10] to show the performance of verification scenarios and compare the widely used systems in the literature of 3D face recognition. For the verification scenario, receiver operating characteristics (ROC) curves for three different FRGC masks, namely ROC I, II, and III are shown in Figure 3. To all of masks at an FAR (False Accept Rate) of 0.1%, our method yields the better verification results through all the thresholds, which corresponds to optimal feature subspaces.

In the Table 1, we demonstrate verification results for ROC I, ROC II, and ROC III protocols, which is treated as the standard evaluation on FRGC v2 3D face recognition. For expression issues, some popular methods is based on the different testing sets, including Neutral vs. Neutral (N vs. N), Non-Neutral vs. Neutral (Non-N vs. N), All vs. Neutral (A vs. N). Table 1 showed the verification results with the FAR of 0.1%. We can conclude that the verification rates

Table 1. Verification results (%) with the different 3D face recognition methods on FRGC v2.0 database

Methods	ROC I	ROC II	ROC III	N vs. N	Non-N vs. N	A vs. N
FRGC [10]	-	-	-	-	40	45
Bettetti et al. [11]	-	-	-	97.7	91.4	95.5
Passalis et al. [12]	-	-	-	94.9	79.4	81.5
Cook et al. [13]	93.71	92.91	92.01	-	-	-
Kakadiaris et al. [14]	97.3	97.2	97	-	-	-
Ours	96.97	96.5	96.1	96.4	92.5	96.3

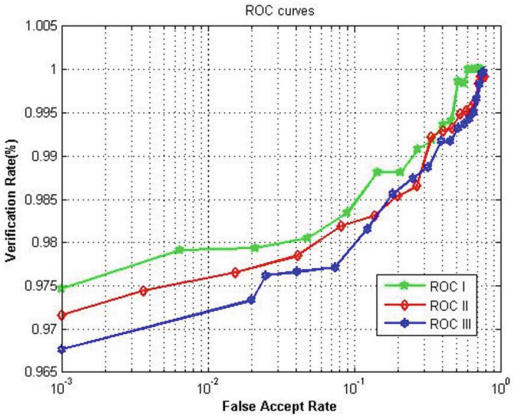

Fig. 3. The ROC curves of FRGC v2.0 3D face database.

of our method are better than other ones, but worse than Kakadiaris et al. [14]. However, Kakadiaris et al. [14] used wavelets and Pyramid transformation to describe the facial scale information and reported the performance was 97% verification at a 0.1% False Acceptance Rate (FAR) in the Face Recognition Grand Challenge. Compared to our algorithm, Kakadiaris et al. [14] utilized a complex algorithm that required a huge computational cost and storage space. For our method, there is a substantial advantage in speed. Our 3D LSIFT algorithm has a much lower computational complexity and is much easier to implement. The performance will be improved when the training data can be significantly increased.

4.2 Performance Evaluation Based on the CASIA 3D Face Database

Here, we employ the CASIA 3D face database [15] and demonstrate the sensor-invariance of our system. This database constitutes challenging variations, including large expressions and poses. The total 4625 facial images can be divided into two subsets, including the training set and the test set. The training set is composed of 615 images, selected 5 facial images for each individual. The rest

Table 2. Rank-1 identification results (%) based on the CASIA 3D face database

Test sets	CO	ADM	SSR	Ours
Illumination Variants	48.21	98.33	96.47	98.5
Expression Variants	45.74	95.73	93.08	95.2
Small Pose Variants	45.27	93.97	92.83	94.86
Large Pose Variants	32.64	56.85	60.99	80.5
Small Pose Variants with Smiling	43.76	90.38	82.15	87.57
Large Pose Variants with Smiling	31.79	52.14	58.43	74.5

of the images from the 123 individuals are used as the test set. In these experiments, we compared the performance of the different popular feature description methods used for 3D face recognition. The considered features include COSMOS shape index (CO) [15], Annotated Deformable Model (ADM) [14], sparse spherical representations (SSR) [16] and our proposed framework for 3D face recognition. The test set is further divided into six subsets to evaluate the performance of different features with pose and expression variations [17]. Table 2 shows the rank-one identification rates for the same test set.

From these results, we can draw the following conclusions: 1) the highest identification rate is up to 98.5% (123 people) obtained by our framework. 2) Shape and illumination variation are important for discriminating an individual. 3) With expression and pose variations, our method and ADM show the better results than other methods. 4) Facial pose variation is another major factor affecting recognition performance. Due to our 3D LSIFT descriptor, our method can capture the shape characteristics of an individual's face and represent this 3D geometric information in an efficient facial regional domain, which shows superior performance relative to other methods that are employed. Thus, our method is robust with the different 3D face database with high generality.

5 Conclusions

A new 3D face recognition method is proposed in this paper by combining a novel facial segmentation method and a novel feature descriptor based on 3D LSIFT. We have utilized a shape bands method for refined region segmentation. Then, we design a unified 3D facial feature descriptor 3D LSIFT fusing appearance and depth information. A huge number of redundant key points can be decreased by adding the facial depth information caused by illumination and pose variations. As a result, our method nicely inherits the invariance of local scale and rotation and increase separability, which overcomes expression and pose variations to some extent. Finally, experimental results demonstrate the improved performance than other popular approaches with a good generalization.

Acknowledgments. The work presented in this paper was supported by the National Natural Science Foundation of China (Grants No. NSFC-61402046 and NSFC-61403024), Fund for the Doctoral Program of Higher Education of China (Grants

No.20120005110002), Beijing Municipal Commission of Education Build Together Project, Principal Fund Project, and President Funding of Beijing University of Posts and Telecommunications.

References

1. Ming, Y., Ruan, Q.: A mandarin edutainment system integrated virtual learning environments. Speech Communication **55**, 71–83 (2013)
2. Bowyer, K.W., Chang, K., Flynn, P.: A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition. Computer Vision and Image Understanding **101**, 1–15 (2006)
3. Bai, X., Li, Q., Latecki, L.J., Liu, W.: Shape band: a deformable object detection approach. In: CVPR 2010, pp. 1335–1342 (2010)
4. Ming, Y., Ruan, Q., Hauptmann, A.: Activity recognition from kinect with 3d local spatio-temporal features. In: ICME 2012, pp. 344–349 (2012)
5. Ming, Y., Ruan, Q.: Robust sparse bounding sphere for 3d face recognition. Image and Vision Computing **30**, 524–534 (2012)
6. Ming, Y., Ruan, Q., Ni, R.: Learning effective features for 3d face recognition. In: ICIP 2010, pp. 2421–2424 (2010)
7. Moreno, A.B., Sanchez, A., Velez, J.F., Diaz, F.J.: Face recognition using 3d surface-extracted descriptors. In: IMVIP 2003 (2003)
8. Alyuz, N., Gokberk, B., Akarun, L.: Regional registration for expression resistant 3d face recognition. IEEE Trans. Information Forensics and Security **5**, 425–440 (2010)
9. Faltemier, T.C., Bowyer, K.W., Flynn, P.J.: A region ensemble for 3d face recognition. IEEE Trans. Information Forensics and Security **3**, 62–73 (2008)
10. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: CVPR 2005, pp. 947–954 (2005)
11. Berretti, S., Bimbo, A.D., Pala, P.: 3d face recognition using isogeodesic stripes. IEEE Trans. Pattern Analysis and Machine Intelligence **32**, 2162–2177 (2010)
12. Passalis, G., Kakadiaris, I.A., Theoharis, T., Toderici, G., Murtuza, N.: Evaluation of 3d face recognition in the presence of facial expressions: an annotated deformable model approach. In: FRG 2005 (2005)
13. Cook, J., McCool, C., Chandran, V., Sridharan, S.: Combined 2d/3d face recognition using log-gabor templates. In: ICVSBS 2006 (2006)
14. Kakadiaris, I.A., Passalis, G., Toderici, G., Murtuza, M.N., Lu, Y., Karampatziakis, N., Theoharis, T.: Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. IEEE Trans. Pattern Analysis and Machine Intelligence **29**, 640–649 (2007)
15. Beumier, C., Acheroy, M.: Automatic 3d face authentication. Image and Vision Computing **18**, 315–321 (2000)
16. Llonch, R.S., Kokiopoulou, E., Tosic, I., Frossard, P.: 3d face recognition with sparse spherical representations. Pattern Recognition **43**, 824–834 (2010)
17. Xu, C., Li, S., Tan, T., Quan, L.: Automatic 3d face recognition from depth and intensity gabor features. Patter recognition **42**, 1895–1905 (2009)