3D Face Recognition Using Local Features Matching on Sphere Depth Representation

Hanchao Wang, Zhichun Mu(✉), Hui Zeng, and Mingming Huang

School of Automation and Electrical Engineering,
University of Science and Technology Beijing, Beijing 100083, China
{wanghanchao314,hzeng,mmhuang1205}@163.com, mu@ies.ustb.edu.cn

Abstract. This paper proposes a 3D face recognition approach using sphere depth image, which is robust to pose variations in unconstrained environments. The input 3D face point clouds is first transformed into sphere depth images, and then represented as a 3DLBP image to enhance the distinctiveness of smooth and similar facial depth images. An improved SIFT algorithm is applied in the following matching process. The improved SIFT algorithm employs the learning to rank approach to select the keypoints with higher stability and repeatability instead of manually rule-based method used by the original SIFT algorithm. The proposed face recognition method is evaluated on CASIA 3D face database. And the experimental results show our approach has superior performance than many existing methods for 3D face recognition and handles pose variations quite well.

Keywords: 3D face recognition · Sphere depth image · Local binary patterns · Learning to rank

1 Introduction

The face has its own advantages over other biometrics for human recognition, since it is natural, nonintrusive, and contactless. Using 3D face scans has been proposed as an alternative solution to conventional 2D face recognition approaches, which is more robust to pose and lighting variations. However, there are still questions worth studying in 3D face recognition in real world application. In this paper, we propose a method of dealing with 3D face recognition in unconstrained scene.

1.1 Related Work

The task of recognizing 3D face scans has been approached in many ways, leading to varying levels of successes [1]. Global-based recognition uses the entire face region to compute similarity. E.g., PCA [2], LDA [3] and ICP-based matching [4]. Facial surface shape descriptors can be used in the face recognition, too. Such as curvature [5], point signature [6], Extended Gaussian Image (EGI) [7] signed shape-difference map [8] and eLBP [9] etc. These holistic methods usually need an accurate normalization step with respect to pose and scale changes.

J. Yang et al. (Eds.): CCBR 2015, LNCS 9428, pp. 27–34, 2015.
DOI: 10.1007/978-3-319-25417-3_4

To improve the face recognition performances of various post changes in unconstraint scenes, some studies using multiple multi-view images to identify human faces [10]. Liu et al [11] use the spherical fitting point cloud data to convert 3D data to 2D depth image. This conversion keeps complete three-dimensional structural information, transforming the post changes to rotation changes.

1.2 Motivation and Approach Overview

In a conventional face recognition experiment, both the probe and gallery scans are acquired cooperatively in a controlled environment so as to precisely capture and represent the whole face [12]. Differently, in semi-cooperative or uncooperative scenarios, probe scans are acquired under unconstrained conditions that may result in face scans with post variations or partial data, thus demanding for methods capable of performing recognition in unconstraint condition.

In this paper, we first establish the spherical representation of the 3D human face using a sphere fitting algorithm. 3D face shape is then represented as sphere depth image. Using 3DLBP Depth Face [9] generated on depth image, we apply SIFT [13] descriptors to the sphere 3DLBP image for the detection of keypoints. In order to extract keypoints with more repeatability and robustness, learning to rank [14] algorithm is adopted to the keypoints extraction. Proposed method is tolerant of face pose variations and gains higher recognition rates when compared with other 3D face recognition methods. The general framework is illustrated in Fig. 1.

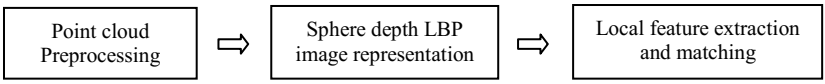

Fig. 1. General framework of proposed method

The paper can be summarized as follows: Section 2 describes the point cloud preprocessing and the sphere depth image generation. In section 3, we introduce the local features matching on sphere LBP depth image utilizing learning to rank strategy. The experimental results are provided in section 4. Section 5 concludes the paper.

2 Generation of Sphere Depth Image

The basic assumption is that the shape of human head can be viewed as a ball. Sphere depth image of 3D face shape is an object-centered shape representation using data-fitting Algorithm. This object-centered representation is uncorrelated with pose variation which is very appropriate in uncooperative scenarios.

A 3D face shape is fitted on a sphere through a linear least square method [11] to get the optimal solution of sphere parameters. After the sphere center and radius is acquired, the sphere depth r is defined as the distance between the point and the sphere surface. Then the Cartesian coordinates of input point clouds can be transformed to spherical coordinates $[r, \theta, \rho]$ .The points in spherical coordinates then can construct a sphere depth image using interpolation on θ and ρ.

This sphere representation completely displays the 3D face shape without any information loss. This representation is invariant of scale, pose, and illumination, which is essential to the face recognition without face-alignment.

3 Local Feature Extraction on Sphere Depth Image

There are some local features remain invariant compared to the corresponding 3D facial scan in the gallery set. Once located and characterized accurately, these local features can be utilized to identify 3D faces.

The 3DLBP [15] on depth images extracts the absolute depth value difference between the central pixel and its neighbors, which is better describe local shapes than origin depth image. Multiple layers of 3DLBP depth images are further used in local features matching process. Combining these local features on multiple layers of LBP depth images can improve the final accuracy.

3.1 Problem in Keypoints Selection

SIFT features [13] is widely used to calculate a similarity score between two object images in the subsequent matching process. However, Faces are non-rigid, round and smooth objects. The depth changes in face images are gradual and slow, the blob and corner structures are not significantly different from their neighboring pixels. In the keypoint detection of SIFT, The candidate keypoints need to be selected by experiential thresholds.

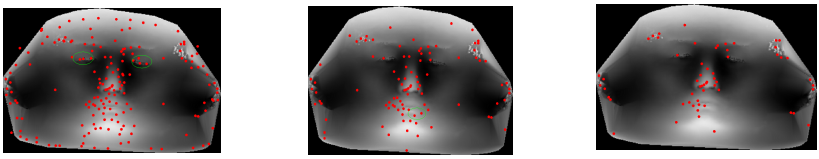

Fig. 2. Keypoints detected in a face image: (a) The initial keypoints; (b) The keypoints remain after removal keypoints with low contrast; (c) The keypoints remain after applying a threshold with high edge responses.

Fig. 2 shows the deficiencies of unreliable keypoints removal algorithm in the SIFT approach; (a) shows the full extracted keypoints; (b) illustrates the keypoints remain after applying a threshold γ_1 on the minimum contrast of each candidate keypoint; (c) shows the final keypoints remain after further removing keypoints with high edge responses γ_2. The keypoints marked with ellipse are eliminated by the removal strategies. But these keypoints can represent distinctive structures, and are in the area of high repeatability with the pose changes.

Therefore, the keypoint removal scheme in the SIFT framework will eliminate some useful features when applied to face images.

3.2 Ranking Keypoints in Keypoints Selection

To overcome the problem above, we adopt a supervised approach to select the keypoints rather than using thresholds. Keypoint repeatability $R(x_i)$ is defined as the times that the same keypoints appear in a sequence of same person's face images. It is a significant measure of the keypoint robustness and stability. We can select keypoints according to their repeatability.

We use the learning to rank algorithm [14] to train a ranking model to rank keypoints according to their stability. The features utilized in the ranking model are associated with the steps in keypoints extraction and removal, including the first/second derivatives of the depth image, the eigenvalues (λ_1, λ_2), determinant Det (H), and the eigenvalue ratio $Trac\ (H)^2/Det(H)$ of the Hessian matrix H. Suppose x_i, x_j are two keypoints in train images, if $R(x_i) > R(x_j)$, we obtain a pair $< x_i > x_j >$.The Rank-SVM train these pairs to find the optimal classification function ,which is also the optimal rank function to rank these keypoints. However, the original Rank-SVM treats the keypoint ranking scores equally, and the original Rank-SVM treats the difference in the pairs equally, which ignores the fact that points with higher repeatability and pairs with more disparity is more concerning in the keypoint matching.

Therefore, we set a weight $Q(x_i, x_j)$ for different pair distinguish their contribution to the rank. The $Q(x_i, x_j) = |R(x_j) - R(x_j)|*[R(x_i) + R(x_j)]$. $|R(x_i) - R(x_j)|$ is the difference of two points' repeatability in pair. $[R(x_i) + R(x_j)]$ denotes the significance of keypoints location in rank list. The larger the $Q(x_i, x_j)$ is, the more crucial the pair is.

We construct a SVM model to solving this Quadratic Optimization problem:

$$\begin{aligned} &\min_{\omega,\xi ij} \frac{1}{2} \| \omega \|^2 + C \sum Q(x_i, x_j) \xi_{ij} \\ &s.t < \omega x_i - x_j > 1 - \xi_{ij}; \forall x_i > x_i, \xi_{ij} \geq 0 \end{aligned} \tag{1}$$

We utilize the weight of SVM function ω to form a ranking function for ranking keypoints.

$$f_{\vec{w}}(\vec{x}) = \langle \vec{w}, \vec{x} \rangle \tag{2}$$

In the following features matching steps, we can use the ranking function to select original keypoints with high repeatability rather than using the threshold to choose keypoints manually.

4 Experiment Analysis

To examine the performance on pose changes and partially occluded 3D facial scans, we compared our method with existing methods using the CASIA 3D face database. This database covers 123 subjects and contains complex variations in pose that are challenging for any algorithm. Fig. 3 shows some example of the database with pose variations. We select front faces of 80 people as the gallery set, and the multi-pose images of these people as the test set.

(a) (b) (c) (d) (e) (f) (g)

Fig. 3. Example of CASIA 3D face: (a) Up 20°-30°. (b) Tilt left 20°-30°. (c) Left 20°-30°. (d) Front (e) Right 20°-30°. (f) Tilt right 20°-30°. (g) Down 20°-30°.

The raw 3D face scans consist not only human heads but also shoulders, necks as well as noise points. We use median filter to smooth the human face surface. The face and non-face area can be separated using thresholding technique. These extracted point clouds of face region then are converted to sphere depth images and sphere LBP images using the method mentioned in section3. Fig. 4 illustrates some sphere depth images and different layers of LBP images.

We apply SIFT to each layer of LBP images and fuse the features in the hybrid matching process, making use of the original weight calculation as proposed in [16] to dynamically determine corresponding weights of each layers. The overall numbers of matching pairs calculated above then are used as the similarity to identify the test image in the gallery.

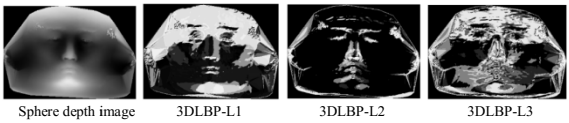

Fig. 4. Sphere depth images and different layers of 3DLBP images

We establish the ranking model in keypoint extraction and matching, a training set is constructed by calculating repeatability of candidate keypoints appearing in an image sequence of different poses as ranking scores, including Left Right, Tilt left, Tilt right, Up and Down. We adopt the ranking model shown in section 3.2.

4.1 Experiment on Ranking Model

To validate the effectiveness of ranking model, we then compare the repeatability of the same number of interest points both extracted by SIFT and improved SIFT algorithm Weight Rank-SIFT. Original SIFT use the standard threshold γ_1=0.03, γ_2=10, while the top ranked interest points obtained by Rank-SIFT and Weight Rank-SIFT methods are kept. As shown in Table 1, the Weight Rank-SIFT algorithm selects keypoints has higher repeatability than the other two methods.

Table 1. Keypoints repeatability in different images with pose changes.

Repeatability	Left 30°	Right 30°	Tilt Left 30°	Tilt Left 30°	Up 30°	Down 30°
SIFT	0.421	0.413	0.476	0.462	0.436	0.418
Rank-SIFT	0.476	0.466	0.512	0.485	0.471	0.462
Weight Rank-SIFT	0.493	0.477	0.541	0.533	0.486	0.467

We further compare the recognition performance between Weight Rank-SIFT, Rank-SIFT and SIFT with same numbers of keypoints under five different parameter configurations. From Fig. 5, we can see that different thresholds result in different performance. Weight Rank-SIFT method have higher recognition rate than the original Rank-SIFT and SIFT using manual threshold to select keypoints.

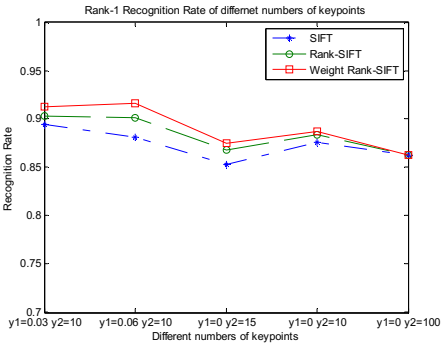

Fig. 5. Rank-1 recognition rate of different numbers of keypoints

4.2 Experiment on Pose Change 3D Faces Images

We compare our approach with some existing methods. We divide the test set into Small Pose Variation (SPV): including views of front, left/right 20–30°, up/down 20–30°and tilt left/right 20–30°. Large Pose Variation (LPV): including views of left/right 50–60°. Table 2 and Fig. 6 shows that our method preserves the stable local features in post invariant spherical representation has better performance in Rank-1 recognition rate even the objects have large pose changes. Although Mesh-SIFT [19] have better performance than our method, it costs more time to process on 3D mesh directly, while our method based on depth image is rapid in real world application. To sum up, the experimental results prove that with selecting repeatable local features from spherical representation, our method can deal with pose changes and even partial missing data in 3D face recognition.

Table 2. Recongnition rate on CASIA pose change testset

Approach	SPV	LPV	Overall
ICP	74.2%	64.7%	67.1%
3DLBP	89.2%	68.4%	82.1%
Depth + Intensity [17]	89.2%	70%	85.2%
RBGR [18]	93.2%	82%	89.6%
SIFT on Sphere 3DLBP	91.4%	84.5%	88.4%
Extended-LBP [9]	N/A	N/A	90.4%
V-LBP[20]	N/A	N/A	90.4%
Mesh SIFT	N/A	N/A	92.1%
Weight Rank-SIFT	93.4%	85.6%	91.4%

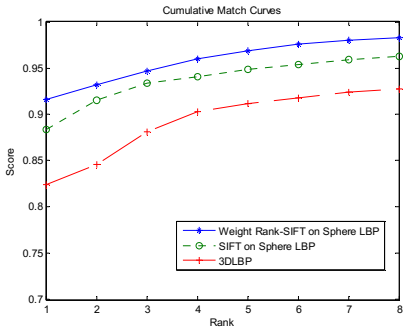

Fig. 6. CMC curves on pose change test set

5 Conclusion

Our method has been proposed to address a challenging topic, 3D face recognition in uncontrolled environments, in which pose variations is a big obstacles. We have presented an effective approach to solve the problem using sphere depth image and 3DLBP representation. This pose-invariant representation allows for accurate and fast description of local shape variations, thus enhancing the distinctiveness of 3D faces. An improved SIFT algorithm Weight Rank-SIFT is then used to select more stable and repeatable keypoints on face image to the following local features matching process. The promising recognition performance shows its potential to deal with the 3D face recognition in unconstraint scene.

Acknowledgements. This paper was supported by (1) National Natural Science Foundation of China under the Grant No. 61472031; (2) Portions of the research in this paper use the CASIA-3D FaceV1 collected by the Chinese Academy of Sciences' Institute of Automation (CASIA).

References

1. Bowyer, K., 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), 1–15 (2006)
2. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience **3**(1), 71–86 (1991)
3. Belhumeur, P., Hespanha, J., Kriegman, J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(7), 711–720 (1997)
4. Lu, X., Jain, A., Colbry, D.: Matching 2.5D face scans to 3D models. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(1), 31–43 (2006)
5. Gordon, G.: Face recognition based on depth and curvature features. In: Proceeding of IEEE Conference Computer Vision and Pattern Recognition, pp. 808–810 (1992)
6. Chua, C., Han, F., Ho, F.: 3D human face recognition using point signature. In: Proceeding International Conference Automatic Face and Gesture Recognition, pp. 233–238 (2000)

7. Tanaka, H., Ikeda, M., Chiaki, H.: Curvature-based face surface recognition using spherical correlation–Principal directions for curved object recognition. In: Proceeding International Conference Automatic Face and Gesture Recognition, pp. 372–377 (1998)
8. Wang, Y., Liu, J., Tang, X.: Robust 3D face recognition by local shape difference boosting. IEEE Transaction Pattern Analysis and Machine Intelligence **32**(10), 1858–1870 (2010)
9. Ouamane, A., Belahcene, M., Bourennane, S.: Multimodal 3D and 2D face authentication approach using extended LBP and statistic local features proposed. In: 4th European Workshop on Visual Information Processing, pp. 130–135 (2013)
10. Zhang, X., Gao, Y.: Face recognition across pose: A review. Pattern Recognition **42**(11), 2876–2896 (2009)
11. Liu, P., Wang, Y., Zhang, Z.: Representing 3D face from point cloud to face-aligned spherical depth map. International Journal of Pattern Recognition and Artificial Intelligence 26(01) (2012)
12. Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: IEEE Workshop on Face Recognition Grand Challenge Experiments, pp. 947–954 (2005)
13. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer and Vision **60**(4), 91–110 (2004)
14. Li, B., Xiao, R., Li, Z.: Rank-SIFT: Learning to rank repeatable local interest points. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1737–1744 (2011)
15. Huang, Y., Wang, Y., Tan, T.: Combining Statistics of Geometrical and Correlative Features for 3D Face Recognition. In: British Machine Vision Conference, pp. 879–888 (2006)
16. Mian, A., Bennamoun, M.: Keypoint detection and local feature matching for textured 3D face recognition. International Journal of Computer and Vision **79**(1), 1–12 (2008)
17. Xu, C., Li, S., Tan, T.: Automatic 3D face recognition from depth and intensity Gabor features. Pattern Recognition **42**(9), 1895–1905 (2009)
18. Ming, Y.: Robust regional bounding spherical descriptor for 3D face recognition and emotion analysis. Image and Vision Computing **35**, 14–22 (2015)
19. Smeets, D., Keustermans, J., Vandermeulen, D.: meshSIFT: Local surface features for 3D face recognition under expression variations and partial data. Computer Vision and Image Understanding **117**(2), 158–169 (2013)
20. Tang, H., Yin, B., Sun, Y.: 3D face recognition using local binary patterns. Signal Processing **93**(8), 2190–2198 (2013)