3D Face Recognition Fusing Spherical Depth Map and Spherical Texture Map

Shuai Liu[1], Zhichun Mu[1(✉)], and Hongbo Huang[2]

[1] School of Automation and Electrical Engineering,
University of Science and Technology Beijing, Beijing 100083, China
ustbliu2013@163.com, mu@ies.ustb.edu.cn
[2] Computing Center, Beijing Information Science
and Technology University, Beijing 100192, China
hauck@sohu.cn

Abstract. Face recognition in unconstrained environments is often influenced by pose variations. And the problem is basically the identification that uses partial data. In this paper, a method fusing structure and texture information is proposed to solve the problem. In the register phase, the approximate 180 degree information of face is acquired, and the data used to identify individual is obtained from a random single view. Pure face is extracted from 3D data first, then convert the original data to the form of spherical depth map (SDM) and spherical texture map (STM), which are invariant to out-plane rotation, subsequently facilitating the successive alignment-free identification that is robust to pose variations. We make identification through sparse representation for its well performance with the two maps. Experiments show that our proposed method gets a high recognition rate with pose and expression variations.

Keywords: Face recognition · Spherical Depth Map · Spherical Texture Map · Sparse representation

1 Introduction

In modern society, person identification is very helpful, especially in some important roles of public information and security. In practical application, face recognition has received much attention in the distance and uncontrolled scene for its friendly and non-intrusive [1]. Although great advances have been made in 2D face recognition, it is still a challenging task to overcome the effect of light and pose [2]. Research has begun to focus on 3D data recent years with the development of 3D scanner. Compared with 2D data, 3D point cloud preserves the structure information. However, it still has limits in occlusion and data missing.

In fact, person identification in unconstrained environments is basically the identification that uses partial data. But, registration is possible under controlled environments, able to acquire data contained the approximate 180 degree information of face. Build a library, including the complete individual information, could help us use the partial data obtained from a random single view to identify individual. As we all

J. Yang et al. (Eds.): CCBR 2015, LNCS 9428, pp. 151–159, 2015.
DOI: 10.1007/978-3-319-25417-3_19

know, ear recognition has become an effective identification method because it is a rigid body, containing rich structure features [3]. And ear data acquisition is similar to face data, both can realize non-intrusive acquisition. In addition, ear has unique physical location, when face to the side, ear is in positive. All the reasons make ear be a useful supplement to face recognition. Some researchers get a panoramic 2D image by stitching multiple angle views of head for face recognition. Liu [4] used a statistical model for face image alignment, and introduced a recognition method by projecting images to the surface of a 3D ellipsoid. Singh [5] proposed a hierarchical registration algorithm to align different view faces and they had experiments on three databases. Fan [6] built a simple acquisition system composed of 5 standard cameras which, together, could take simultaneously 5 views of a face at different angles. Then they chose an easily hardware-achievable algorithm, consisting of successive linear transformations, to compose a panoramic face. At last, recognition experiments based on principal component method was conducted. Rama [7] presented a method for the automatic creation of 180° aligned cylindrical projected face images using nine different views. The alignment was done by applying first a global 2D affine transformation of the image, and afterward a local transformation of the desired face features using a triangle mesh. This local alignment allows a closer look to the feature properties and not the differences. Finally, these aligned face images were used for training a pose invariant face recognition approach.

But, most 3D face recognition is based on structure information only, and texture features are used for 2D face recognition [8, 9]. However, texture and structure infor-mation are two different description forms, which have a certain complementarity [10].

In this paper, we propose a method fusing 3D depth data and 2D texture data. Pure head is extracted first because point cloud obtained through 3D scanner contains not only face and ear, but also includes a shoulder and other redundant information. Then, a sphere is fitted to the 3D data, based on the fact that human head looks like a ball. 2D spherical depth map (SDM) is generated by expanding the fitting sphere and the pixel value on the map is the distance of point cloud to the center of the fitting sphere. Meanwhile, according to the correspondence between the original texture and depth information, texture is mapped to the fitting sphere and we expand the sphere like before to generate the spherical texture map (STM). At last, we make identification through sparse representation using the two maps.

The rest of this paper is organized as follows: Section 2 presents a method to extract pure face from 3D point cloud by detecting nose tip and central profile. In Section 3, the recognition method using spherical depth map and spherical texture map is illustrated in detail. The experiments and results are provided in Section 4. Finally, some concluding remarks are given in Section 5.

2 Pure Face Extraction

Point cloud obtained through 3D scanner contains not only face and ear, but also includes a shoulder and other redundant information. It will have a significant impact later whether we get rid of the redundant information accurately. Considering the fact that nose tip and central profile are usually easy to detect in a random single view within the scope of 180 degree of face, we extract pure face from 3D point cloud by detecting the obvious part.

About nose tip detection, one simple method is to assume that it is the point closest to the camera, but it will fail under non-frontal pose. However, the tip still has the maximum depth value as long as the original coordinate rotating to the right direction, shown in Fig. 1.

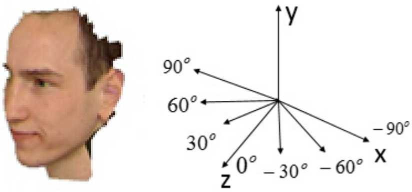


Fig. 1. Rotated Coordinate

So, nose tip will have the largest local polar radius after cylindrical coordinate transformation. Let θ and R denote the rotated angle and polar radius in cylindrical coordinate system respectively, and y is the point cloud vertical coordinate with the same meaning as in Cartesian coordinates.

We first assume that every point may be nose tip. For a point P, search its neighboring points P_i within the distance of 10. If one polar radius of P_i is larger than that of P, we no longer believe P is a nose tip. Certainly, this method has a large computational load on localizing the neighboring points. Considering the actual situation, we don't consider the points with their y too large or too small, and the points whose polar radius less than 40% of the largest polar radius are also ignored. The rest points are calculated by KD-Tree algorithm. The final candidate nose tip points are marked in Fig. 2.

For further conform, we calculate the Shape Index of every candidate point and some points around it. The calculation is in original point cloud data. The Shape Index was put forward by Dorai [11], computed as:

$$ShapeIndex(s) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{k_1(s)+k_2(s)}{k_1(s)-k_2(s)}. \tag{1}$$

where $k_1(s)$ and $k_2(s)$ represent the maximum and minimum principal curvature respectively. In our experiments, the Shape Index is set to 0.7-1.

After nose tip P was confirmed, take out all the points in the box, which width is 30 and height is 25, shown in Fig. 3. And y-coordinate of the points in the box is quantized to an integer. Let F_i be the set $f=(f.\theta, f.y, f.R)$ such that $|f.y|=i$, the point with the largest polar radius of F_i belongs to nose ridge. The θ and y belong to nose ridge is denoted by a matrix $A_{m\times 2}$ and the θ and y of all pure face points is denoted by a matrix $A'_{m\times 2}$, where m is the points number of nose ridge and n is the points number of pure face. PCA is used to A, able to get two feature vectors $v1$, $v2$ and two characteristic values $r1$, $r2$ ($r1>r2$). Let $\theta'=A\times v2$, the points in pure face whose θ' is equal to θ'_P are the points belong to central profile, shown in Fig. 4. Where θ'_P is the θ' of nose tip P.

Central profile bottom could be confirmed because nose tip is almost the midpoint of the central profile, and pure face is extracted by getting rid of the points below the bottom point.

Fig. 2. Candidate nose tips in cylindrical coordinate

Fig. 3. Nose ridge in cylindrical coordinate

Fig. 4. Central profile in Cartesian coordinate

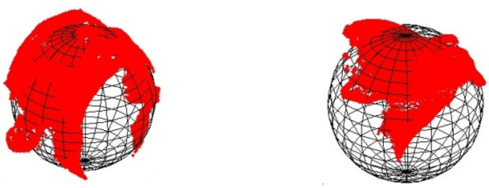
3 Recognition Process

3.1 Spherical Depth Map and Spherical Texture Map

A sphere is fitted to 3D face point cloud since the shape of human head can be approximated by a sphere. The spherical equation in 3D Cartesian coordinates is:

$$(x-x_0)^2+(y-y_0)^2+(z-z_0)^2=r^2. \tag{2}$$

The Spherical Depth Map (SDM) was first introduced in [12]. In [12], a linear least square method is used to solve the fitting problem. Fig. 5 shows the fitting result.



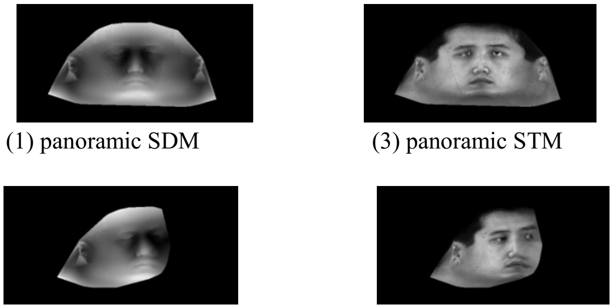
(1) complete data (2) partial data

Fig. 5. Sphere fitting

After moving the original coordinate to the center of fitting sphere, Cartesian coordinate is translated to spherical coordinate. The theta and phi which are inclination angle and azimuth angle in spherical coordinate, make up the coordinates of spherical depth map. Fig. 6 (1), (2) shows the panoramic SDM and partial SDM generated by holistic data and partial data. The pixel on the map is the distance of point cloud to the center of fitting sphere. Because point cloud is discrete, linear interpolation is needed.

Texture information is mapped to the fitting sphere as the original texture information has correspondence with the original depth information. Then spherical texture map (STM) can be obtained like spherical depth map, shown in Fig. 6 (3), (4).

Panoramic SDM and STM can be obtained in the register phase, holistic structural and textural information are made available, which is definitely helpful to alleviate the problems induced by pose variations and facial expression. Partial SDM and STM produced by test data acquired from a random single view. These spherical maps, in addition, are invariant to out-plane rotation, subsequently facilitating the successive alignment-free identification that is robust to pose variations. Meanwhile, the way of data presenting of the maps in 2D form reduces the cost of data storage and the load of computation involved in recognition process.



(1) panoramic SDM (3) panoramic STM

(2) partial SDM (4) partial STM

Fig. 6. SDM and STM

3.2 Sparse Representation

Person identification in unconstrained environments is basically the identification that uses partial data of the 180 degree face. So we can't use classification method based on global features for identification due to the different feature numbers among the partial images. Sparse representation classification method is applied to image classification and recognition widely because its state-of-the-art performance for image block or data missing [13]. Liao [14] developed an alignment-free face representation method based on Multi-Keypoint Descriptors (MKD), where the descriptor size of a face is determined by the actual content of the image. In that way, any probe face image, holistic or partial, can be sparsely represented by a large dictionary of gallery descriptors. In this paper, we investigate the depth image and texture image with multi-task sparse representation classification method respectively, the feature used for the two maps is Affine Sift [15].

Assume we have m features for each holistic spherical image i, corresponding descriptors are $D_i=[d_{i1}, d_{i2}, ..., d_{im}]$, and denote by D the collection of n images:

$$D = \left(D[1], D[2], \cdots, D[n]\right). \tag{3}$$

where D is a dictionary containing $m = \sum_{i=1}^{n} m_i$ descriptors.

Given a test example $Y=(d_{y1}, d_{y2}, ..., d_{yK})$, which belongs to one of the n classes, our goal is to find the class to which the test example belongs. We are interested in solving the following optimization:

$$X = \arg\min_{X} \sum_{k=1}^{K} \|x_k\|_0, \quad \text{s.t. } Y = DX . \tag{4}$$

where $X=[x_1, x_2, \ldots., x_K]$ are sparse coefficient and $\|\cdot\|_0$ denotes the l_0 semi-norm indicating the number of nonzero elements of the given vector, that is $\|x\|_0=\sum_k I(x_k \neq 0)$, where $I=1$ is true. Since the P_{l0} optimization program is NP-hard, a convex relaxation of it is obtained by replacing the l_0 with the l_1 norm, which is $\|x\|_1=\sum_k|x_k|$, and solving the following convex program:

$$X = \arg\min_{X} \sum_{k=1}^{K} \|x_k\|_1, \quad \text{s.t. } Y = DX . \tag{5}$$

In this paper, we convert the problem to K minimization problems:

$$\begin{aligned} \hat{x}_k &= \arg\min_{x_k} \|x_k\|_1, \\ s.t.\ \ y_k &= Dx_k, k = 1, 2, \cdots, K \end{aligned} \quad . \tag{6}$$

And the Eq. (6) is solved by the Homotopy method [16]. Then reconstruction error for every class is calculated:

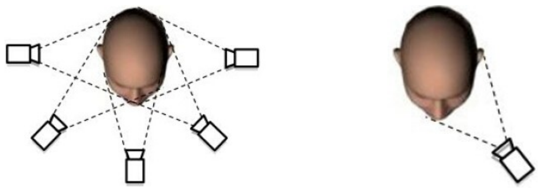
$$r_i(Y) = \frac{1}{K} \sum_{k=1}^{K} \|y_k - D[i]\hat{x}_k[i]\|_2^2 \cdot \tag{7}$$

where $x[i]$ denotes that we select coefficients related with class i only, and others are set to 0. Besides, the negative coefficients are also set to 0 for the non-negativity is more consistent with the biological modeling of visual data [17]. The test example Y belongs to the class with the smallest $r(Y)$.

4 Experiments

4.1 Database

Since there is no existing public library containing the complete data, a small set containing 50 people is built for our experiments. The complete individual data is obtained though the multiple perspectives and we fuse the data by ICP algorithm [18]. Meanwhile, the test data is obtained from a random single view and part of the test data contained expression varieties. Every single view is taken two times, and 1700 images in total. Fig. 7 shows the schematic diagram of data acquired for registration and test.

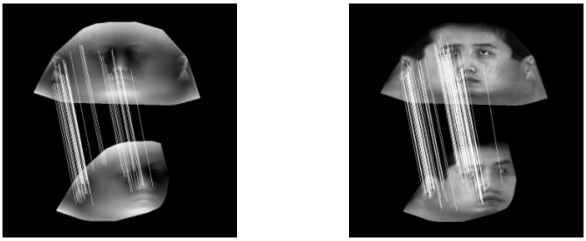


(1) Registration (2) Test

Fig. 7. Data acquisition for registration and test

4.2 Recognition

Features are extracted from spherical depth map and spherical texture map respectively, using Affine Sift. And the parameter transition tills defined in Affine Sift is set to 3. The data used in our experiments are acquired under different posture, containing the rotation of three axes. We first identify individuals through sparse representation using SDM and STM respectively, and the results are shown in Table 1, 2. The match features are shown in Fig. 8.



(1) SDM Match (2) STM Match

Fig. 8. Match features

And then make identification fusing the two maps. We assume Eq. (8) the probability of the test sample belongs to class i, $r(i)$ denotes the reconstruction error. According to Bayesian decision, the test sample belongs to the class with maximal probability P, which is the product of P' calculated by SDM and STM. And the results are shown in Table 3.

$$P' = \frac{1/r_i^2}{\sum_{j=1}^{n} 1/r_j^2} . \tag{8}$$

Table 1. SDM

α \ β	0	±45	±90
0	93.3%	86.5%	93.5%
30	91.0%	84.5%	92.5%
-30	88.0%	85.0%	94.0%

Table 2. SDM + STM

α \ β	0	±45	±90
0	96.7%	90.0%	95.5%
30	96.0%	89.5%	94.5%
-30	94.0%	91.5%	95.0%

Table 3. STM

α \ β	0	±45	±90
0	95.3%	88.5%	92.0%
30	93.0%	85.5%	89.5%
-30	90.0%	83.0%	90.5%

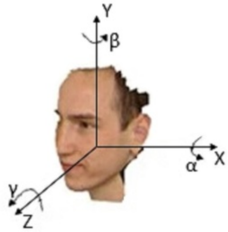


Fig. 9. Rotation of three axes

As can be seen, we can identify individual using the information acquired from a random single view. And if we fuse the texture and structure information together, a better result can be obtained.

5 Conclusion

In this paper, we propose a 3D face recognition method fusing spherical depth map and spherical texture map. The 3D data of pure face is converted to 2D data, and at the same time we save the structure information and texture information completely. After panoramic registration in the case of controlled circumstances, we can recognize individual using the partial data acquired from a random single view. This will have a high application value for person identification in unconstrained environments.

Acknowledgments. This paper is supported by (1) National Natural Science Foundation of China under the Grant No. 61472031; (2) National Natural Science Foundation of China under the Grant No. 61170116.

References

1. Jain, A.K., Flynn, P., Ross, A.A. (eds.).: Handbook of biometrics, Springer, US (2007)
2. Zhang, H., Zhang, Y., Huang, T.S.: Pose-robust face recognition via sparse representation. Pattern Recognition **46**(5), 1511–1521 (2013)

3. Nixon, M.S., Bouchrika, I., Arbab-Zavar, B., Carter, J.N.: On use of biometrics in forensics: Gait and ear. In: European Signal Processing Conference (2010)
4. Liu, X., Chen, T.: Pose-robust face recognition using geometry assisted probabilistic modeling. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1:502–509. IEEE (2005)
5. Singh, R., Vatsa, M., Ross, A.: A mosaicking scheme for pose-invariant face recognition. IEEE Transactions Systems, Man, and Cybernetics, Part B: Cybernetics **37**(5), 1212–1225 (2007)
6. Yang, F., Paindavoine, M., Abdi, H., Monopoli, A.: Development of a fast panoramic face mosaicking and recognition system. Optical Engineering, 44(8) (2005)
7. Rama, A., Tarres, F., Rurainsky, J.: Aligned texture map creation for pose invariant face recognition. Multimedia Tools and Applications **49**(3), 545–565 (2010)
8. Lei, Y., Bennamoun, M., El-Sallam, A.A.: An efficient 3D face recognition approach based on the fusion of novel local low-level features. Pattern Recognition **46**(1), 24–37 (2013)
9. Mohammadzade, H., Hatzinakos, D.: Iterative Closest Normal Point for 3D Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(2), 381–397 (2013)
10. Huang, D., Ardabiliab, M., Wang, Y., Chen, L.: Asymmetric 3D/2D face recognition based on LBP facial representation and canonical correlation analysis. In: 16th IEEE International Conference on Image Processing (ICIP), pp.3325–3328. IEEE (2009)
11. Dorai, C., Jain, A.K.: COSMOS-A representation scheme for 3d free-form objects. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(10), 1115–1130 (1997)
12. 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(1) (2012)
13. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(2), 210–227 (2009)
14. Liao, S., Jain, A.K., Li, S.Z.: Partial face recognition: Alignment-free approach. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(5), 1193–1205 (2013)
15. Yu, G., Morel, J.M.: ASIFT: An Algorithm for Fully Affine Invariant Comparison, Image Processing On Line. 1 (2011)
16. Donoho, D.L., Tsaig, Y.: Fast Solution of L1-Norm Minimization Problems When the Solution May be Sparse. IEEE Transactions on Information Theory **54**(11), 4789–4812 (2008)
17. Zhang, B., Mu, Z., Li, C., Zeng, H.: Robust classification for occluded ear via Gabor scale feature-based non-negative sparse representation. Optical Engineering, 53(6) (2014)
18. Besl, P.J., Mckay, N.D.: A method for registration of 3-D shape. IEEE Transaction on Pattern Analysis and Machine Intelligence **14**(2), 239–256 (1992)