3D Face and Ear Recognition based on Partial MARS Map

Tingting Zhang, Zhichun Mu, Yihang Li, Qing Liu and Yi Zhang
School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China
926zhangtingting@sina.com, mu@ies.ustb.edu.cn, {liyihang_1992, ustbliuqing}@163.com

Keywords: 3D Face Recognition, Partial MARS Map, Deep Learning, Head Pose Estimation.

Abstract: This paper proposes a 3D face recognition approach based on facial pose estimation, which is robust to large pose variations in the unconstrained scene. Deep learning method is used to facial pose estimation, and the generation of partial MARS (Multimodal fAce and eaR Spherical) map reduces the probability of feature points appearing in the deformed region. Then we extract the features from the depth and texture maps. Finally, the matching scores from two types of maps should be calculated by Bayes decision to generate the final result. In the large pose variations, the recognition rate of the method in this paper is 94.6%. The experimental results show that our approach has superior performance than the existing methods used on the MARS map, and has potential to deal with 3D face recognition in unconstrained scene.

1 INTRODUCTION

Face recognition is one of the most popular biometric feature recognition methods with the characteristics of nonintrusive, contactless, accessible and informative. However, the single features from human face image are susceptible to the variations of age, expression and pose. Ear recognition is another biometric recognition method with the advantages of its insusceptible to the age and expression variations. A multi-biometric feature fusion and recognition method based on face and ear has been proposed, which is more robust to pose and expression issues. However, there are still some questions to be solved of 3D face recognition in unconstrained scene. In this paper, we propose a method based on partial MARS map, which represents the better recognition performance of dealing with large pose variations in the unconstrained scene.

Many face recognition algorithms have been proposed over the decades which based on global features and local features. The first methods require normalization of the images lacking of robustness, and the second methods always neglect the global information. Then a new method of fusing the global and local features is presented (Melzer et al., 2003). In order to improve the recognition result in the unconstrained scene, some studies take the use of multi-view 3D point clouds to identify human faces (Zhang and Gao, 2009). Considering the positional relationship of face and ear, some researchers (Liu et al., 2015; Wang et al., 2015; Huang, 2015) propose a new method based on MARS depth map and texture map. They merge 3D face point clouds from the multi-view without any influence factors to a new point cloud with more information including the complete face and ear. Then they transform the expression of Cartesian coordinates into spherical coordinates, and generate the MARS depth and texture maps. The sphere depth map (SDM) and sphere texture map (STM) of the candidate 3D face point clouds are also required to be matched with MARS depth maps and MARS texture maps of gallery set respectively. This method presents a better recognition performance compared with other general methods in the case of complete ear and incomplete face. However, in other unconstrained conditions, especially existing large pose variations, it doesn't work well.

In the case of large pose variations, SDM and STM always present a certain degree of deformation, which will increase the recognition difficulty when matching with the entire MARS depth and texture maps without any deformation. This paper proposes a new method based on partial MARS map to deal with large pose variations. In this paper, we estimate the facial pose situations based on deep learning, and then generate the partial MARS maps corresponding to each facial pose situation, then generate the 3DLBP (Huang et al., 2006) feature maps from MARS map, and choose

Zhang, T., Mu, Z., Li, Y., Liu, Q. and Zhang, Y.
3D Face and Ear Recognition based on Partial MARS Map.
DOI: 10.5220/0006244206330637
In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)*, pages 633-637
ISBN: 978-989-758-222-6

the sift feature descriptor at the stage of feature extraction. In the recognition stage, we adopt an improved method from sparse representation classification (Liao et al., 2013) to get the matching result, and then use the Bayesian fusion method (Chen et al., 2014) between the matching results from SDM and STM.

The paper is organized as follows: Section 2 describes the pose estimation using deep learning and the generation of partial MARS map. Section 3 presents the feature extraction based on 3DLBP feature maps and the process of recognition. The experimental results are provided in Section 4. Finally, some concluding remarks are given in Section 5.

2 DEEP LEARNING AND MARS MAP

In this paper, we use a deep learning method (Xu et al., 2015) with convolutional neural network (CNN) for facial pose estimation to classify the images from the different databases into 5 poses. The detailed network structure is designed as follows: There are three convolutional layers, two max-pooling layers, a fully connected layer and the soft-max output layer indicating five classes which stand for the five poses. The first two convolution layers are weight sharing and have 64 convolution kernels with size of 5*5 respectively. The following two max-pooling layers also have 64 convolution kernels, and the size of each kernel is 2*2. The third convolutional layer is fully connected to the second max-pooling layer, and the last hidden layer is fully connected to the third convolutional layer. In the soft-max output layer with 5 classes, the class with the maximum probability is expected the estimated one.

In the image preprocessing stage, we take two steps: Firstly, we extract the face region from original images; then we crop each processed face image into five patches, and resize patches into 32*32 as the input of the network. The patches correspond to the four corners and central part with eighty percent of the whole image. In the training stage, we use total five patches as training data, and in the testing stage, we only use the central patch for estimation.

2.1 Facial Pose Estimation using Deep Learning

In this paper, we use a deep learning method (Xu et al., 2015) with convolutional neural network (CNN) for facial pose estimation to classify the images from the different databases into 5 poses. The detailed network structure is designed as follows: There are three convolutional layers, two max-pooling layers, a fully connected layer and the soft-max output layer indicating five classes which stand for the five poses. The first two convolution layers are weight sharing and have 64 convolution kernels with size of 5*5 respectively. The following two max-pooling layers also have 64 convolution kernels, and the size of each kernel is 2*2. The third convolutional layer is fully connected to the second max-pooling layer, and the last hidden layer is fully connected to the third convolutional layer. In the soft-max output layer with 5 classes, the class with the maximum probability is expected the estimated one.

In the image preprocessing stage, we take two steps: Firstly, we extract the face region from original images; then we crop each processed face image into five patches, and resize patches into 32*32 as the input of the network. The patches correspond to the four corners and central part with eighty percent of the whole image. In the training stage, we use total five patches as training data, and in the testing stage, we only use the central patch for estimation.

Our algorithm is tested in CMU PIE database and CAS PEAL face database. The CMU PIE database contains 68 subjects with 13 poses. We choose 19584 images with the 5 poses mentioned above. The CAS PEAL face database contains 1040 subjects with 21 poses, and we choose 4160 images from this database including the images with 5 poses in the Pose database and all the images in the Normal database. Of the total data, the first 50 persons from the CMU PIE database and the data from the Normal database in the CAS PEAL face database are selected to train the network, and the rest of data is used for testing. The network with a learning rate of 10-3 is obtained by the 300 epochs. In our experiment, we get the accuracy of 98.7% in the training stage, and 98.4% in the testing stage.

2.2 Generation of Partial MARS Map

We select the effective area of the entire MARS map according to each situation of the 5 views and generate 5 partial MARS maps. Firstly, 3D face images are collected from the 3D scanning optical system, and each subject contains three point clouds respectively from three views of the left, right and front side. After merging the three point clouds, we get a more complete 3D point cloud called the fusion

point cloud. The coordinate distribution is shown as Figure 1.

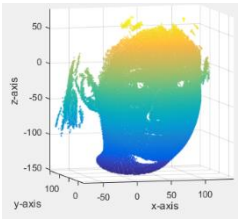

Figure 1: The mosaic of three 3D face point cloud.

We fit the 3D face shape on a sphere model according to the linear least square method to get the sphere center point C and the sphere radius Radii. Then determine a axis through point C with the normal vector (0,0,1), and rotate the 3D face point cloud to 90°, 45°, 0°, -45°, -90°. After rotation of the point cloud, we select the effective region as followed: Firstly, we divide the 3D point cloud into three parts along the z axis equally, which are upper, middle, bottom part. Then determine the boundary point of the middle part and acquire the value of the variable y as Yt. The boundary point is defined as the edge point of the ear in the opposite side to the rotation direction. We assume that the point with the maximum or the minimum value of variable x in the middle part is the boundary point. When the rotation direction is left, we get the point with the minimum of x, and if the rotation direction is right, we get the point with maximum of x. Finally, we select the region with the value of y smaller than Yt as the valid region. Then the Cartesian coordinates [x, y, z] of the new rotated point cloud can be transformed to spherical coordinates [r, θ, ρ]. The partial MARS Depth maps with five views are displayed as Figure 2 for examples. And we can see that the partial MARS map does not mean the half of the entire original MARS map in the case of 90° pose variation.

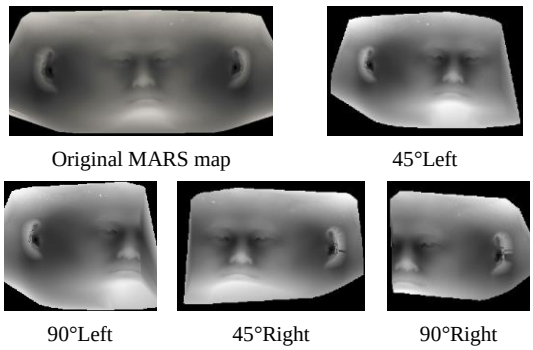

Figure 2: The original MARS depth map and partial MARS depth maps.

3 RECOGNITION PROCESS

3.1 Local Feature Extraction on SDM and STM

After generation of all partial MARS maps, we do the same processing with the candidate 3D point clouds to get SDM and STM. In the feature extraction stage, an improved algorithm called Weight Rank-SIFT (Wang et al., 2015) is used to select more stable feature points. However, it is difficult to extract enough feature points from SDM. The 3DLBP (Huang et al., 2006) maps generated from the SDM can extract the absolute depth value difference between the central pixel and its neighbors and presents a better representation regarding feature description. The combination of the local features on SDM and the first two layers of LBP maps can improve the recognition rate. Figure 3 shows the representation of key points extracted from SDM and the first three layers of 3DLBP maps. With regards to STM, we extract features from the STM directly.

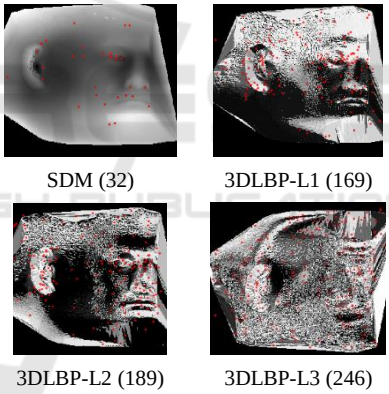

Figure 3: Features extracted from SDM and the first three layers of 3DLBP maps.

3.2 Decision Fusion based on Sparse Representation Classification

The SDM and STM in the probe set are the representation of the candidate 3D point clouds, and the number of features on them is different from the number of features from partial MARS maps. In this paper, Multi Key points Descriptor Sparse representation Classification (MKD-SRC) is used to classification. In this method, any candidate SDM and STM in the probe set can be sparsely represented by a dictionary consisted of the descriptors in the gallery set. In this paper, we calculate the average residual of the candidate point

cloud descriptor in the dictionary. The smaller average residual is, the higher matching degree is. We use the MKD-SRC method on SDM and STM respectively, and then fuse the average residual scores of them by a Bayesian method (Chen et al., 2014) to get the final matching score to determine which subject the candidate 3D face image belongs to.

4 EXPERIMENTAL ANALYSIS

To examine the performance on large pose variations, we build a small set containing 20 people with 90 3D face images of each person. The images include complex variations with one or two factors in pose, occlusion and expression. Considering the shortage of data, we selected the 3D face images of 80 subjects from the CASIA 3D face database, and get the MARS maps through the ICP algorithms. Finally, our experiment database contains 20 subjects collected from our own equipment and 80 subjects from CASIA 3D face database, and the images of each person contain the normal state without any influence factors and the states with factors of 5 pose variations, laughter, smile, angry, sad and close eyes which are similar with the uncontrolled conditions. We select one MARS depth map and one MARS texture map of each subject as the gallery set, and select the SDMs and STMs generated by the candidate images of these subjects as the test set. In this experiment, we compare our approach with some existing methods used on the MARS map based on the test set and the results are shown in Table 1 (the recognition rate of 90° is the average value of the +90° and -90°). In general, our method has a slight advantage with 90°, and has better performance with 45°. Towards the frontal view without any influence factors, the recognition rates of four methods are close to each other. From the Table 2, we can conclude that in the cases of angry and close eyes with the 45° view of pose variations, our method presents a satisfied recognition rate. But in the laughter cases, eyes and mouth of the person always have the large deformations, and the performance still need to be improved.

Table 1: Recognition rate in the cases of single pose variation issue.

Approach	90°	45°	0°	Overall
Huang	93.5%	N/A	95.7%	94.6%
SDM+STM	95.5%	90.0%	96.7%	94.1%
3DLBP+SDM	93.8%	85.6%	95.3%	91.6%
Our Approach	95.7%	91.3%	96.7%	94.6%

Table 2: Recognition rate in the cases of expressions with 45°rotation.

Rotation angle	laughter	angry	close eyes
0°	75.5%	96.3%	87.0%
45°	83.8%	91.4%	89.3%
90°	95.2%	95.3%	94.3%

5 CONCLUSIONS

Our approach has been proposed for the unconstrained scene, in which pose and expression variations are the difficult issues to be solved. This paper has presented an effective method to solve the large pose variations issue by facial pose estimation and feature matching on partial MARS map. 3DLBP representation is used to acquire the more feature points on SDM and partial MARS depth map. After feature extraction by Weight Rank-SIFT, we choose the MKD-SRC method during the matching process, and then perform the fusion operation at the decision level. The experimental results show that this method has well promising potential to be applied in the uncontrolled environments.

ACKNOWLEDGEMENTS

This paper was supported by National Nature Science Foundation of China under the Grant No.61472031.

REFERENCES

Melzer T, Reiter M and Bischof H. (2003) “Appearance models based on kernel canonical correlation analysis”, Pattern Recognition, Vol. 36, pp. 1961-1971.

Xiaozheng Zhang, Yongsheng Gao. (2009) “Face recognition across pose: A review”, Pattern Recognition, Vol. 42, pp. 2876-2896.

Shuai Liu, Zhichun Mu and Hongbo Huang. (2015) “3D Face Recognition Fusing Spherical Depth Map and Spherical Texture Map”, Biometric Recognition.

Hanchao Wang, Zhichun Mu, and Hui Zeng. (2015) “3D Face Recognition Using Local Features Matching on Sphere Depth Representation”, Biometric Recognition.

Hongbo Huang. (2015) “Multi-modal Recognition of Face and Ear based on MARS Map”, University of Science and Technology Beijing.

Yonggang Huang, Yunhong Wang and Tieniu Tan. (2006) "Combining Statistics of Geometrical and Correlative Features For 3D Face Recognition", British Machine Vision Conference.

Shengcai Liao, Anil K Jain and Stan Z Li. (2013) "Partial Face Recognition: Alignment-Free Approach", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, pp. 1193-1205.

Long Chen, Zhichun Mu and Baoqing Zhang. (2014) "Ear recognition from one sample per person", Plos One, Vol. 10.

Xiao Xu, Lifang Wu, Ke Wang, Yunkun Ma and Wei Qi. (2015) "A Facial Pose Estimation Algorithm Using Deep Learning", Biometirc Recognition.