A 3D Face Recognition Method Using Region-Based Extended Local Binary Pattern

Shiwen Lv, Feipeng Da, Xing Deng
School of Automation, Southeast University, Nanjing 210096, China
Key Laboratory of Measurement and Control of CSE, Ministry of Education

ABSTRACT

A 3D face recognition method using region-based extended local binary pattern (eLBP) is proposed. First, the depth image converted from the preprocessed 3D pointclouds is normalized. Then, different regions according to their distortions under facial expressions are extracted by binary masks and represented by the uniform pattern of extended LBP. Finally, sparse representation classifier (SRC) is adopted for classification on the single region. Feature-level and score-level fusion with weight-sparse representation classifier (W-SRC) are also tested and compared, and the latter has better performance. The experiments on FRGC v2.0 database demonstrate that the proposed method is robust and efficient.

Index Terms**—**3D face recognition, extended local binary pattern, depth image, binary mask, weight-sparse representation classifier

1. INTRODUCTION

In the past twenty years, 2D face recognition has been one of the most important and attractive topics in computer vision and pattern recognition [1]. However, illumination and pose variances have affected the recognition performance seriously. With these inherent disadvantages and limitations of 2D face data, 3D face data has attracted more interest. Compared with the 2D data, the 3D face data includes more spatial information, which can lead to better recognition performance. 3D face recognition algorithms can be classified into holistic-based, region-based and hybrid-based algorithms. Holistic-based algorithms focus on extracting global features of 3D face data. Pan *et al*[2] maked face points projection into a 2D isomorphic plane. Then the corresponding mapping depth image was established in this plane and subspace was used for matching strategy. However, variations in pose and expression can influence the recognition performance. Region-based algorithms extract region or local feature and match them for recognition. For example, Berretti *et al*[3]extracted 85 landmarks of face and calculated the sift description of these landmarks. Finally, descriptors were matched by the SVM classifier. Li *et al*[4] matched the central profile by LTS-HD to eliminate a large number of candidate face. At last, the extended ICP algorithm was used as a measure to match the rigid region. Hybrid-based algorithms which fuse local feature and global feature are proposed to improve the recognition performance. While Main[5] revealed that the weights of the global and local feature were not easy to decide.

The algorithms based on region focus on the characteristics in different regions. Compared with the holistic-based algorithms, the region-based ones have the advantages to deal with illumination and pose variances. A new 3D face recognition method using region-based extended LBP is proposed in this paper.

The remainder of this paper is organized as followed. The preprocessing is introduced in Section 2. Section 3 presents the feature extraction. W-SRC is introduced in Section 4. Section 5 discusses the experimental results and analysis. Finally, Section 6 concludes this paper.

2. PREPROCESSING

2.1. 3D Face preprocessing

The original pointclouds from FRGC v2.0[6] have spikes and holes. In our previous work[7], we presented a method for the 3D face preprocessing. The method is used in this paper and some important details are introduced as followed. Before converting pointclouds to depth image, the spikes are removed by the Gaussian filter and holes are filled by the bicubic interpolation.

Nose tip that is a convex of the face can be determined according to the shape index[8] and geometric constraint. Apart from the region of face, the original pointclouds has some unnecessary parts, for example ears and shoulders. These unnecessary parts affect the face recognition performance. To extract the feature of face, a sphere of radius r=90mm centered at nose tip is then used to crop the 3D face. The original data and cropped face are shown in Fig.1.

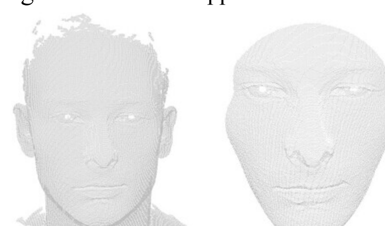


Fig.1 (a) Original face and (b) cropped face

Face alignment is important for the SRC. PCA is used to normalize the face pose so that all the faces can be translated to frontal pose. The new coordinate system is called pose coordinate system (PCS) in which the origin is the nose tip. The Y axis is the eigenvector corresponding to the largest eigenvalue and Z axis is the eigenvector of the smallest eigenvalue. After translating every face into the PCS, all the faces are normalized with the preciously processing method.

2.2. Depth image and normalization

In depth image, the 3D pointclouds is represented by the gray image. The pixel value of every point indicates the distance between the actual point and the camera. In a sense, the depth image is similar to the 2D image. So many mature 2D face recognition algorithms can be applied into the depth image. The depth images are computed by interpolating at the integer x and y coordinates along the horizontal and vertical index respectively and determining the corresponding z coordinate as a pixel value. The pixels in the depth image are then re-sampled at a distance of 1 mm along both the x and y directions. Finally, all the depth images are normalized into a certain size 200*200 to build a public coordinate system in which all the nose tips' coordinate are (100,100) (see Fig.2).

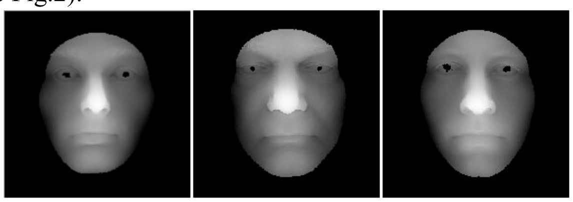


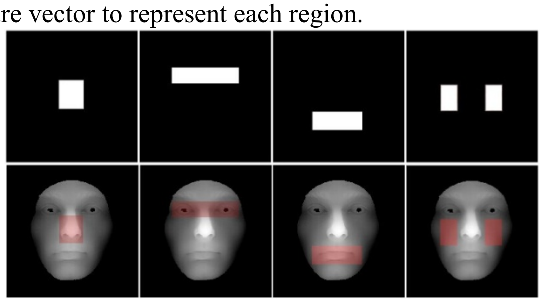
Fig.2 Depth image of face after normalization

3. FEATURE EXTRACTION

The facial expression variation has potential influence on the face recognition due to non-rigid deformation. These deformations may enlarge the differences within class leading to the wrong recognition. The face recognition under expression has become grown a lot of interest in the past decade. According to the distortions by expression, this paper extracts the rigid region, semi-rigid region and non-rigid region of face. And these local regions are represented by the extended LBP.

3.1. Local Region Extraction

According to the expression's distortions, the whole face can be divided into several parts as followed: rigid region(nose), semi-rigid regions (eye-forehead and cheek) and non-rigid region (mouth). Four binary masks are designed to crop the region as shown in Fig.3. The 1st row corresponds to the nose (N), eye-forehead (E), mouth (M) and cheek (C). The eLBP feature extracted from these regions is used as the feature vector to represent each region.



A: nose B: eye-forehead C: mouth D: cheek

Fig.3 The 1st row: binary masks of local region

The 2nd row: the extracted regions from preprocessed face

3.2. Extended LBP

A new local feature proposed in this paper is extended from the LBP[9]. Firstly for a 3*3 neighborhood, according to the gray level values of the central pixel and its surrounding pixel, the Laplace operators of central pixel and its surrounding are calculated. Laplacian operator is an isotropic differential operator with rotation invariance. The Laplace operator of a 2D image function is defined as followed:

$$\nabla^2 f(x,y) = \frac{\partial^2 f}{\partial^2 x} + \frac{\partial^2 f}{\partial^2 y} \quad (1)$$

To deal with better 2D digital image, the discrete expression is shown as Equation (2):

$$\nabla^2 f(x,y) = \left[f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) - 4f(x,y) \right] \quad (2)$$

Then the eLBP operator of the central pixel (x_c, y_c) is calculated as Equation (3) :

$$eLBP(x_c, y_c) = \sum_{n=0}^{7} s(l_n - l_c) 2^n \quad (3)$$

where n covers the eight neighbors of the central pixel, l_c and $l_n (n = 0,1...7)$ are Laplace operators of the central pixel and its surrounding pixel respectively. Function $s(x)$ is compute as follows:

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases} \quad (4)$$

Fig.4 gives an example of this process. Meanwhile the patterns of binary decrease from 256 to 59 the same as the uniform of the LBP [10].

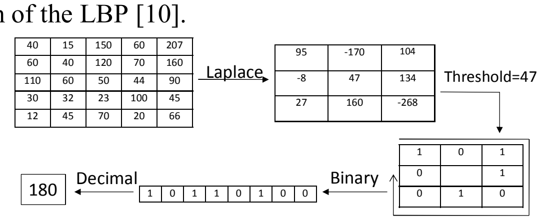


Fig.4 Calculation of eLBP operator

4. W-SRC

4.1. SRC

Wright *et al*[11] introduced the sparse representation classifier into 2D face recognition. The fundamental assumption is that well-aligned frontal face image in different lighting conditions and various facial expression, approximately lies on a linear subspace spanned by sufficient training samples from the same subject. So for a frontal test sample, it can be approximately represented by all the training samples. And if the representation is the most reasonable, the coefficient vector is sparse whose entries are equal to zero or near zero except for the ones associated with the same class. This assumption is also applied into the 3D face recognition. Given N_i training samples of the *i*th class:

$A_i = \left[v_{i,1}, v_{i,2}, v_{i,3} \cdots v_{i,N_i} \right] \in R^{M \times N_i}$, where $v_{ij} \in R^M$ is the *j*th sample of *i*th class. Any test sample $y \in R^M$ of the *i*th class can be represented as follows:

$$y = a_{i,1} v_{i,1} + a_{i,2} v_{i,2} + a_{i,3} v_{i,3} \cdots + a_{i,N_i} v_{i,N_i} \tag{5}$$

where $a_{i,j} \in R, j = 1, 2 \dots N_i$. Theoretically, the test face *y* from *i*th class can be represented by all the training faces $A: y = Ax$. (6)

where $x = \left[0 \cdots 0 \;\; a_{i,1} \;\; a_{i,2} \cdots a_{i,N_i} \;\; 0 \; \cdots \; 0 \right]^T \in R^N$.

Notice that the vector *x* is sparse, so the Equation (6) can be approximately solved by the following problem:

$$\hat{x}_0 = \arg\min \|x\|_0 \;\; s.t. \;\; \|Ax - y\|_2 \le \varepsilon \tag{7}$$

where $\varepsilon \in R^N$ is a deviation vector. Wright[10] pointed that when x_0 is sparse enough, this equation can be solved approximately by solving the problem of L1-Norm:

$$\hat{x}_1 = \arg\min \|x\|_1 \;\; s.t. \;\; \|Ax - y\|_2 \le \varepsilon \tag{8}$$

and computing the residuals:

$$r_i(y) = \left\| y - A\delta_i\left(\hat{x}_1\right) \right\|_2^2 \tag{9}$$

The index of minimal delivers the identity of the probe *y*.

4.2. W-SRC

The distortions under expression variations differ in different regions. So the region having better performance is given higher weight. The feature vector of every face combing different regions can be written as $v_i = \left[v_{i1}{}^T \;\; v_{i2}{}^T \;\; \cdots \;\; v_{iL}{}^T \right]^T$, where $v_{ij} \in R^m$ is the *j*th region feature vector of *i*th face and L is the number of local region extracting from the face. Accordingly, the tested face is written as $y = \left[y_1{}^T \;\; y_2{}^T \;\; \cdots \;\; y_L{}^T \right]^T$ and $A = \left[A_1^T \;\; A_2^T \;\; \cdots A_L{}^T \right]^T$ where $A_i = \left[v_{1,i}, v_{2,i}, v_{3,i} \cdots v_{N,i} \right], i = 1, 2 \dots L$. In this case, the equation is equal to solving the Equation (10) :

$$\hat{x}_0 = \arg\min \|x\|_0 \;\; s.t. \sum_{l=1}^{L} w_l \; \|y_l - A_l x\|_2 \le \varepsilon \tag{10}$$

where w_l is the weight of the *i*th region. This is called the weighted sparse representation-based classifier (W-SRC) [12]. The according residuals are followed:

$$r_i(y) = \sum_{l=1}^{L} w_l \left\| y_l - A_l \delta_i\left(\hat{x}\right) \right\|_2^2, i = 1, 2 \dots N \tag{11}$$

If we write

$A_w = \left[w_1 A_1^T \;\; w_2 A_2^T \;\; \cdots \;\; w_L A_L{}^T \right]^T$, $y_w = \left[w_1 y_1^T \;\; w_2 y_2^T \;\; \cdots w_L y_L{}^T \right]^T$.

The Equation (11) is equal to solving

$$\hat{x}_0 = \arg\min \|x\|_0 \;\; s.t. \;\; \|A_w x - y_w\|_2 \le \varepsilon \tag{12}$$

Equation (12) means that the weighted sparse representation model in Equation (10) is equal to solving a single SRC with global feature vectors. The weights of different regions depend on recognition performance of single local region and they satisfy $\sum_{l=1}^{L} w_l = 1$. The weights are calculated in Equation (13):

$$w_l = \frac{R1_l}{\sum_{l=1}^{L} R1_l} \tag{13}$$

where $R1_l$ is the Rank-1 recognition rate of region l.

5. EXPERIMENTAL RESULTS AND ANALYSIS

5.1. FRGC

FRGC v2.0 is one of the largest public 3D face database in the literature and has become the benchmark for evaluating 3D face recognition approaches. The data is made up of 4007 textured 3D face scans of 466 subjects. Every subject has about 22 faces with pose variations and hair occlusion. 60% of the faces are under neutral facial expression and others include laugh, surprise, angry, afraid and other expressions.

5.2. Recognition Performance of Single Region

In this section, the recognition experiment is done in the single region by extracting the eLBP operator. All the faces of the FRGC v2.0 are included not only the neutral ones but also faces with different expressions. The 3D face scan with a neutral expression from each subject makes up a gallery of 466 samples. The dictionary *A* of the SRC is consisted by the eLBP of these faces. Then remaining 3541 faces are treated as probes y. The results are shown in Table.1

Table.1 Classification results of single region

Single Region	Rank-1 RR
N (rigid)	92.03%
E (semi-rigid)	89.83%
M (non-rigid)	43.68%
C (semi-rigid)	87.36%

Notice that the recognition performance of rigid region of nose is better than the semi-rigid regions not only eye-forehead but also cheek. However, the performance including non-rigid region (mouth) is not ideal.

5.3. Feature-level fusion

The whole face cannot be represented by the single region. The distortions by the expression variations differ in different regions. In this section the feature-level fusion of different regions is conducted. For the feature $P = \left[p_1 \;\; p_2 \;\; \cdots p_N \right]^T$ and $Q = \left[q_1 \;\; q_2 \;\; \cdots \;\; q_M \right]^T$ of region A and region B respectively, the new feature vector of feature-level fusion is represented by $\left[p_1 \;\; p_2 \cdots p_N \;\; q_1 \;\; q_2 \cdots q_M \right]^T$. The result of feature-level fusion is shown in Table.2

Notice that the fusion of rigid region and semi-rigid regions enhances the performance and is better than single region. While results of fusion including the mouth indicate the non-rigid region weakens the performance.

5.4. Score-level fusion

The score-level fusion of different regions is also conducted with weights calculated by Equation (13) and W-SRC introduced in section 4.2. The results are shown in Table.2.

Table.2 Classification results of fusion

Fusion	Feature-level	Score-level
N+E	94.77%	95.31%
N+C	93.67%	94.49%
N+C+E	95.88%	97.80%
N+C+E+M	91.47%	92.29%

Compared with the feature-level and score-level fusion of rigid region and semi-rigid region respectively. We can see that the recognition rate of score-level fusion has a 1.24% degree higher than the one of feature-level fusion. From the CMC curve (shown in Fig.5) of fusion with nose, eye-forehead and cheek, we know that score-level fusion is better than the feature-level fusion.

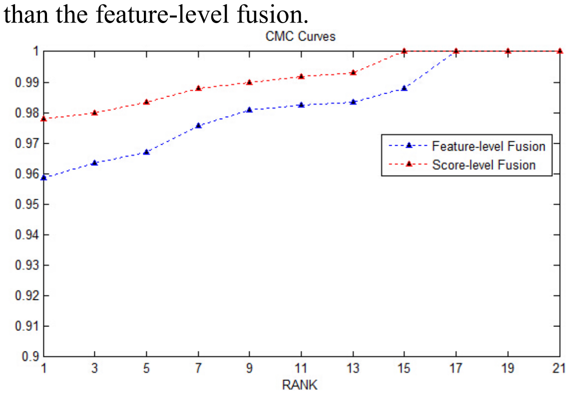


Fig.5 CMC curves of feature-level and score-level fusion

5.5. Result Analysis and Comparison

In this paper, the classification experiments are conducted in the single region including the rigid region, semi-rigid regions and non-rigid region. Feature-level and score-level fusion are also tested. The experiments demonstrate that expression variations have great effect on the non-rigid region. On the contrary, the rigid region is robust to the expression with good recognition performance. Finally, the feature-level and score-level fusion improve effectively the recognition performance. Moreover, the score-level fusion is better than the feature-level fusion.

5.5.1. Comparison to the algorithms based on region

To demonstrate the effectiveness of algorithm proposed in this paper, the comparison to results published on the FRGC v2.0 data set is shown in Table.3. All these algorithms are based on extracting the features from the local regions of face.

Table.3 Comparison of results from different algorithms based on region

Method	Rank-1 RR
Main *et al*[13]	97.37%
Lei *et al*[14]	95.60%
Faltemier *et al*[15]	97.20%
Al-Osaimi *et al*[16]	97.40%
This paper	97.80%

5.5.2. Comparison to feature based on LBP

To demonstrate the effectiveness of the eLBP operator proposed in this paper, the comparison to extended LBP operators published on the FRGC v2 data set is shown in Table.4.

The same as this paper, the pointclouds are translated into depth image in literature [17] and [18]. Then LBP features are extracted to conduct the recognition experiments. The recognition performance of literature [18] is similar to this paper. While, it combines the extended LBP and multi scale sift descriptor. So it is more complex.

The literature [19] divides the pointclouds into blocks. The depth and normal information are represented by the LBP operator and the score-level fusion is also tested. While the LBP feature extracted from the 3D model is easily influenced by the noise. The comparison from Table.4 demonstrates the effectiveness of the eLBP descriptor in local region in this paper.

Table.4 Comparison of results from different algorithms based on LBP

Feature	Rank-1 RR
MS-ELBP-DFS[17]	97.20%
DLBP[18]	90.00%
V-LBP[19]	94.89%
This paper	97.80%

6. CONCLUSION

A 3D face recognition algorithm using region-based extended local binary pattern is proposed. Firstly, 3D pointclouds is preprocessed and PCS is used to put all the faces into the same frontal pose. Then, the normalized depth image of all the faces is obtained. Moreover, different regions according to their distortions under facial expressions are extracted by binary masks and represented by the uniform pattern of extended LBP. Finally, classification experiments by SRC are conducted on the single region. Feature-level and score-level fusion with W-SRC are also tested and compared. The better performance is obtained in latter fusion. The experiments on FRGC v2.0 database demonstrate that the proposed method is robust and efficient.

The following conclusions can be drawn:

1. The algorithm is robust to the expression through fusion of rigid and semi-rigid regions.
2. A new extended LBP feature is proposed and it effectively represents the local regions of face.
3. The score-level fusion with W-SRC enhances the recognition performance and it is better than the feature-level fusion.

7. REFERENCES

[1]Abate F. Abate, Michele Nappi, Daniel Riccio, Gabriele Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters*, pp.1885–1906, vol.28, no.14, 2007.

[2]Gang Pan, Shi Han, Z.H. Wu, Y.M.Wang, "3D face recognition using mapping depth images," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, pp.1-7, 2005.

[3]S.Berretti, A.D. Bimbo, P.Pala, "A set of selected sift features for 3d facial expression recognition," *Proceedings of the 20th International Conference on Pattern Recognition*. Istanbul, pp. 4125-4128, 2010.

[4]L.X. Li, F.P. Da, "A rapid method for 3D face recognition based on rejection algorithm," *Acta Automatica Sinica*, pp.153-158, vol.36, no.1, 2010.

[5]F.R. Al-Osaimi, M. Bennamoun, A. Mian, "Integration of local and global geometrical cues for 3D face recognition," *Pattern Recognition*, pp.1030–1040, vol.41, no.3, 2008

[6]P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, "Overview of the face recognition grand challenge," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, pp. 947-954, 2005.

[7]Li X, Da F, "Efficient 3D face recognition handling facial expression and hair occlusion," *Image and Vision Computing*, pp. 668-679, vol.20, no.9, 2012.

[8]Chitra Dorai, Anil K. Jain, "COSMOS-A representation scheme for 3D free-form objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1115-1130, vol.19, no.10, 1997.

[9]D. Huang, C.F. Shan, M. Ardabilian, Y.H. Wang, "Local Binary Patterns and Its Application to Facial Image Analysis: A Survey," *IEEE Transactions on Systems, Man, and Cybernetics—Part c: Applications and Reviews*, pp.765-780, vol.41, no.6, 2011.

[10]Timo Ojala., Matti Pietikainen, David Harwood, "Multiresotion gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on Pattern Analysis and Machine Intelligenc, pp. 971-987, vol.24, no.7, 2002.

[11]Wright J, Yang A Y, Ganesh A, Sastry S S, Ma Y, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.210–227,vol.21, no.3, 2009.

[12]H.B. Li, Di Huang, J.M. Morvan, L.M. Chen, "Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns," *Neurocomputing*, pp.179-193, vol.133, 2014.

[13]Ajmal S. Mian, M. Bennamoun, Robyn Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.29, no.11, 2007.

[14]Y.J. Lei, M. Bennamoun, Amar A. El-Sallam, "An efficient 3D face recognition approach based on the fusion of novel local low-level features," *Pattern Recognition*, 2013.

[15]T.C. Faltemier, K.W. Bowyer, P.J. Flynn, " A region ensemble for 3-d face recognition," *IEEE Transactions on Information Forensics and Security*, pp.62-72, vol.3, no.1, 2008.

[16]Faisal R. Al-Osaimi, Mohammed Bennamoun, Ajmal Mian, "Spatially Optimized Data-Level Fusion of Texture and Shape for Face Recognition," *IEEE Transactions on Image Processing*, vol.21, no.2, 2012.

[17]D. Huang, M. Ardabilian, Y.H. Wang, L.M. Chen, "A novel geometric facial representation based on multi-scale extended local binary patterns," *Proceedings of IEEE Computer Society Conference on Automatic Face Gesture Recognition and Workshops*. Santa Barbara, CA, pp.1-7, 2011.

[18]Aissaoui, Jean Martinet, Chaabane Djeraba, "DLBP : a novel descriptor for depth image based face recognition," *Proceedings of IEEE International Conference on Image Processing,* Paris, France, october, 2014.

[19]H.L. Tang, B.C. Yin, Y.F. Sun, Y.L. Hu. "3D face recognition using local binary patterns," *Signal Processing*, vol.93, no.8, 2012

8. ACKNOWLEDGMENTS

This research is supported by National Natural Science Foundation of China (No. 51175081, No. 51475092, No. 61405034), Doctoral Fund of Ministry of Education of China (No. 20130092110027).