[1]Jinwei Wang
[2,*]Quan Li

3D Facial Expression Recognition Based on Geometric Feature Fusion

Journal of Electrical Systems

Abstract: - In order to overcome the problems of missing important information and di-mensionality disaster of three-dimensional (3D) facial expression, a novel method based on geometric feature fusion is proposed. This approach initially extracts distance feature vectors and angle feature vectors from key facial regions such as the eyes, mouth, and eyebrows using direct geometric feature. Subsequently, the K-nearest neighbor (K-NN) algorithm is used to obtain the distance feature vector candidate sets and angle feature vector candidate sets separately. Finally, the maximum-minimum rule is utilized to fuse these candidate sets into a single feature vector, completing the recognition of 3D facial expression. Experimental results on the BU-3DFE database demonstrate that this method achieves on overall recognition rate of 97.6%, exhibiting excellent robustness to varia-tions in facial expressions. Furthermore, this approach can serve as a valuable reference for future research in the field of 3D face recognition.

Keywords: 3D facial expression recognition; direct geometric feature; K-nearest neighbor; candi-date set.

I. INTRODUCTION

Facial expressions, as a fundamental means of human emotion expression, serve as an effective tool in interpersonal communication. People can recognize each other’s emotions and inner world through the subtle changes in facial expressions, and also can speculate on each other’s psychology through the changes in facial expressions. The main research object of facial expression recognition is two-dimensional images, however, the face itself is a non-rigid three-dimensional geometric body. Regarding the face as a two-dimensional image will inevitably lose some important information (such as depth information), thus it cannot accurately reflect the subtle changes in facial expressions. Moreover, it is susceptible to factors such as face pose and illumination, while three-dimensional facial data is inherent information of the face, which has good robustness to changes in external conditions. Additionally, three-dimensional facial expression images can be rotated and translated in three-dimensional space at will, which can make up for the differences observed in three-dimensional facial expression images from different visual angles.

With the development of 3D data acquisition technology, scholars are increasingly studying 3D facial expression recognition. Although some good results have been obtained, there are difficulties such as face occlusion and pose variation, especially the elastic deformation of non-rigid regions of the face caused by expression changes[1]. Proposed solutions include methods based on depth, shape model, global, and local features[2], among which the last has attracted much attention because it can effectively solve the problem of expression changes and has good robustness to face occlusion and pose variation. In 2016, Hariri et al. proposed to extract keypoints and corresponding surface patches on 3D faces, and used the constructed local covariance matrix[3]. The method had good recognition rates for both expression and pose variation, but the constructed local covariance matrix had high requirements for linear correlation. In 2017, Guo et al. used value contour lines and mean curvature to detect keypoints, and matched and recognized faces using constructed local features[4]. The method had good recognition rates and robustness for expression variation, but it was time-consuming. In 2017, Emambakhsh and Evans identified the keypoints in the region of the nose tip, using the nose tip itself as a reference, and recognized faces by calculating the histograms of normal vectors of keypoints (used as features)[5]. Recognition rates were high for neutral expressions, but were not ideal when there were significant expression variations. In 2018, Li et al. proposed a face matching and recognition method utilizing multiple specific curve features that can represent facial expressions, and uses the nearest point iteration algorithm[6]. The method had good recognition rates and robustness for expression variation, but it required significant preprocessing work for poses and was time-consuming. In 2020, Zhang et al. proposed a similarity matching method for face recognition by constructing local descriptors using keypoints in rigid regions, and extracting four equidistant geodesic rings in non-rigid regions. The matching results of the two regions were weighted and fused for face recognition[7].

[1] Liming Vocational University, Quanzhou, Fujian, China
[2] Hunan Automotive Engineering Vocational College, Zhuzhou,Hunan, China
*Corresponding author: Quan Li

The method had good recognition rates and robustness for expression variation, but it required recognition in unobstructed conditions.

To solve the problem of missing 3D point cloud data and long identification time， a 3D facial expression recognition based on geometric feature fusion was proposed. The method extracts the distance feature vectors and includes angle feature vectors of the key regions of the eyes, mouth, and eyebrows according to direct geometric features, uses the K-nearest neighbor (K-NN) algorithm to obtain the distance feature vector candidate set and the included angle feature vector candidate set, fuses them as a feature vector, and completes 3D facial expression identification with maximum-minimum rules. The method flow of this paper is shown in Figure 1.

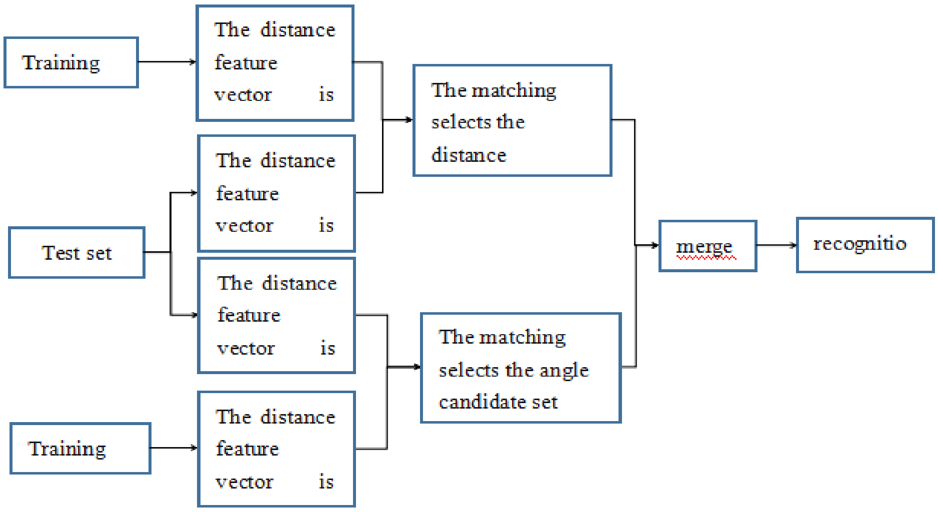
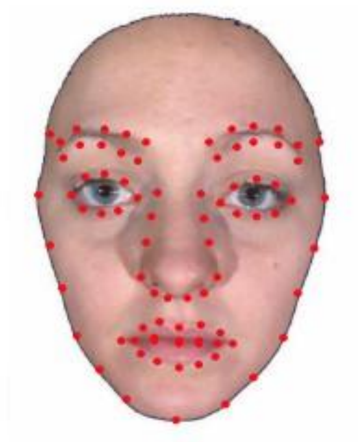


Figure 1 Textual method flow char.

II. Visual Feature Vector Extraction

In the BU-3DFE database, each object model has 83 manually labeled feature points, as shown in Figure 2, which describe facial expression changes using Euclidean distance as visual features.



	#points
right eye	8
left eye	8
right eyebrow	10
left eyebrow	10
nose	12
mouth	20
face boundary	15

(a) (b)

Figure 2 Manual marking of 83 feature points.

(a) The 83 Key Points of 3D Facial Neutral Expression; (b) Distribution of manually labeled points in different regions of face.

Let β_i be a feature point on a three-dimensional facial expression. then the normalized Euclidean distance between two points can be expressed as

$$d_{i,j} = \frac{\|\beta_i - \beta_j\|}{w}, (i < j, i = 1,2,....,83, j = 2,3,....,83) \quad (1)$$

where w is the Euclidean distance between the inner corners of the left and right eyes. Euclidean distance candidate set features for each object model are

$$V_k = (d_{1,2}, d_{1,3}, ..., d_{82,83})^T_{d\times 1}, \tag{2}$$

where $d = C_{83}^2 = 3403$.

In order to reduce the dimensionality of its features, while also improving the difference between different facial expressions.Direct geometric features (DGFs) are used to describe the Euclidean distances and angles between points, including distance feature vectors and angle feature vectors for three key regions: eyes, mouth, and eyebrows.

A. Distance feature vector

DGF describes the Euclidean distance between each pair of feature points, as shown in Table 1. $d_{a,b}$ represents the Euclidean distance between points a and b in a 3D face. Figure 3 shows the order of feature point annotations: points 1–8 are the contour of the right eye, points 9–16 are the contour of the left eye, points 17–26 are the upper and lower points of the right eyebrow, points 27–36 are the upper and lower points of the left eyebrow, points 37–48 are the contour points of the nose, points 49–68 are the inner and outer contour points of the mouth, and points 69–83 are the contour points of the face.

Table 1. Description of DGF distance features.

DGF Feature Type	Feature Name	Representation
Distance feature	Open eyes	$d_1 = d_{11,15}, d_2 = d_{3,7}$
	Tighten brow	$d_3 = d_{27,18}$
	Shift mouth	$d_4 = d_{9,55}, d_5 = d_{1,49}$
	Open mouth	$d_6 = d_{49,55}, d_7 = d_{52,58}$

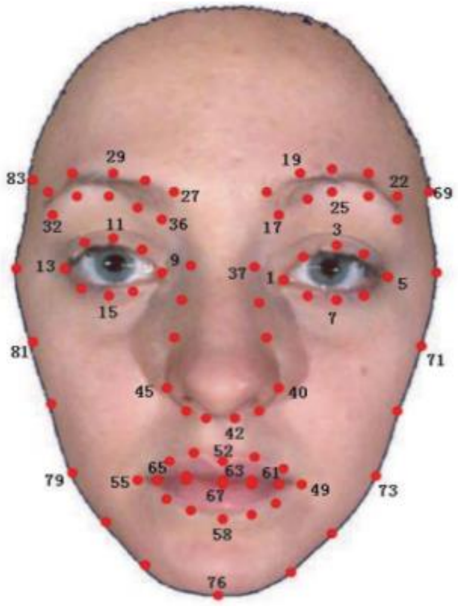


Figure 3. Feature point labeling order.

As shown in Table 1 and Figure 4, a normalized Euclidean distance is used to form a distance feature vector:

$$d = [d_1, d_2, d_3, d_4, d_5, d_6, d_7]. \tag{3}$$

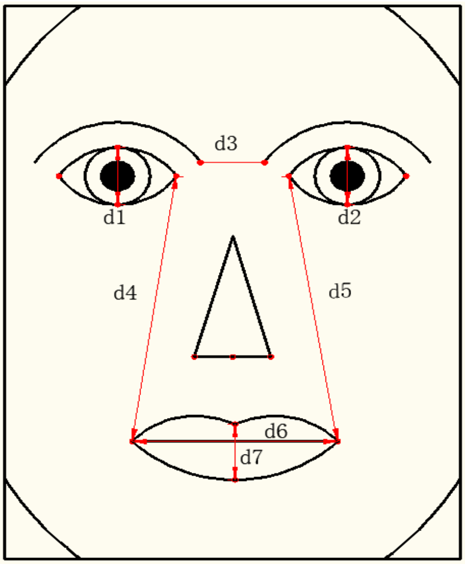


Figure 4 Facial expression distance feature

B. Angle feature vector

The distance feature vector describes the changes in the location of facial expressions, but cannot describe changes in facial expression shapes. Therefore, the DGF angle feature[8] is introduced, as shown in Table 2.

Table 2. DGF angle features.

Feature Type	Feature Name	Representation
Included angle feature	Open eyes between angles	$\theta_1,\theta_2,\theta_3,\theta_4$
	Open mouth at an angle	θ_5,θ_6
	Eyebrow angle	θ_7,θ_8

θ is the angle between vectors $\mathbf{W_{a,b}}$ and $\mathbf{W_{a,c}}$, where $\mathbf{W_{a,b}}$ is a linear vector between spatial feature points a and b, i.e., $W_{a,b}=(x_a,y_a,z_a)-(x_b,y_b,z_b)$.

$\mathbf{W_{a,c}}$ represents the line between spatial feature points a and c, i.e., $W_{a,c}=(x_a,y_a,z_a)-(x_c,y_c,z_c)$, and

$$\theta=\arccos\left(\frac{W_{a,b}\bullet W_{a,c}}{\|W_{a,b}\|\|W_{a,c}\|}\right), \tag{4}$$

$$\theta_1=\arccos\left(\frac{W_{11,13}\bullet W_{11,9}}{\|W_{11,13}\|\|W_{11,9}\|}\right), \tag{5}$$

$$\theta_2=\arccos\left(\frac{W_{13,11}\bullet W_{13,15}}{\|W_{13,11}\|\|W_{13,15}\|}\right), \tag{6}$$

$$\theta_3=\arccos\left(\frac{W_{3,1}\bullet W_{3,5}}{\|W_{3,1}\|\|W_{3,5}\|}\right), \tag{7}$$

$$\theta_4=\arccos\left(\frac{W_{1,3}\bullet W_{1,7}}{\|W_{1,3}\|\|W_{1,7}\|}\right), \tag{8}$$

$$\theta_5=\arccos\left(\frac{W_{52,49}\bullet W_{52,55}}{\|W_{52,49}\|\|W_{52,55}\|}\right), \tag{9}$$

$$\theta_6=\arccos\left(\frac{W_{55,52}\bullet W_{55,58}}{\|W_{55,52}\|\|W_{55,58}\|}\right), \tag{10}$$

$$\theta_7=\arccos\left(\frac{W_{27,31}\bullet W_{27,9}}{\|W_{27,31}\|\|W_{27,9}\|}\right), \tag{11}$$

$$\theta_8 = \arccos(\frac{W_{18,22} \bullet W_{18,1}}{\|W_{18,22}\|\|W_{18,1}\|}) \cdot \tag{12}$$

The above eight angles are formed into the angle feature vector,

$$\theta = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8]. \tag{13}$$

III. 3D FACIAL EXPRESSION RECOGNITION

A. Candidate set of feature vectors

To reduce the accidental misjudgment of expression recognition caused by a single expression image, this paper builds upon the nearest neighbor (NN) algorithm and incorporates the K-nearest neighbor (K-NN) algorithm[9], which searches for the k nearest facial expression images in the training set based on visual feature vectors. This method has good reliability in finding multiple images. The Euclidean distance is calculated as[10],

$$D(q,n) = [\sum_{j=1}^{m}\left(q_j - n_j\right)2]^{1/2}, \tag{14}$$

where q_j represents the j th feature in the distance feature vector for the unknown three-dimensional face image q in the test set, n_j represents the j th feature in the distance feature vector for the 3D face image n in the training set, and $m=7$ is the number of features in the distance feature vector.

We find the k nearest distance feature vectors to form a candidate set of three-dimensional facial images,

$$D = (D_1, ..., D_k) \cdot \tag{15}$$

Similarly, using the cosine angle similarity method, we find the k facial expression images with the highest number of matching points to form a candidate set of angle feature vectors,

$$A = (\alpha_1, ..., \alpha_k), \tag{16}$$

in this case using cosine angle similarity[11],

$$\alpha = \cos^{-1}\frac{\left\langle\theta_q, \theta_n\right\rangle}{\left(\|\theta_q\|\|\theta_n\|\right)}, \tag{17}$$

where θ_q and θ_n are the respective angle feature vectors of the three-dimensional facial images of the test set and training set.

B. Feature vector candidate set fusion

The extracted distance feature vector candidate set D and angle feature vector candidate set A are fused. D and A are normalized separately. The normalized distance feature vector D' and angle feature vector A' are calculated as,

$$D' = \frac{D - \min\{D\}}{\max\{D\}\text{-}\min\{D\}}, \tag{18}$$

$$A' = \frac{A - \min\{A\}}{\max\{A\}\text{-}\min\{A\}} \cdot \tag{19}$$

Finally, the normalized D' and A' are fused using the maximum-minimum principle, i.e.,

$$\boldsymbol{S} = (1\text{-}D') + A' \cdot \tag{20}$$

The vector $\boldsymbol{S}$ value serves as a measure of the similarity between an unfamiliar three-dimensional facial expression image in the test set and the facial expression images present in the training set. A smaller $\boldsymbol{S}$ value indicates greater similarity.

The specific steps involved in the recognition of three-dimensional facial expressions are as follows:

(1) Extract the Euclidean distance vector and angle vector of the unknown three-dimensional facial expression images in the test set;

(2) The K-NN algorithm is used to find the k nearest images to the unknown three-dimensional facial expression images and the k images with the smallest angles in the training set, forming respective candidate sets D and A. They are then fused into a single vector $\boldsymbol{S}$ using the maximum-minimum principle;

(3) A smaller $\boldsymbol{S}$ indicates a higher similarity between an unfamiliar three-dimensional facial expression image in the test set and the facial expression images present in the training set , completing the expression recognition. Figure 1 describes the recognition process.

IV. Experimental Results and Analysis

Experiments were conducted using MATLAB software on a computer with an i5 CPU and 16 GB of memory using the BU-3DFE 3D facial expression database, which includes 100 objects, with 56 being females and 44 being males, Each object is composed of 6 facial expressions and 1 neutral expression, with the 6 facial expressions including 4 intensity variations and 1 neutral expression without any variation. As shown in Figure 5, there are 25 facial images for each subject, for a total of 2500 images.

NE AN DI FE HA SA SU

Figure 5 BU-3DFE database example.

40 subjects (20 females and 20 males) were selected for the experiments, each with 25

facial models; hence $N = 40 \times 25 = 1000$. The models were divided into 10 sets, with two sets randomly chosen as the test set, and the remaining eight used as the training set. The mean value of the 5 experiments was utilized as the experimental outcome. The K-NN algorithm was used in this study, and the value of k affected the recognition rate of 3D expressions. Therefore, experiments were conducted with different values of k, as shown in Table 3. The highest recognition rate was achieved when $k = 5$.

Table 3. The recognition results of different k values.

k	2	3	4	5	6	7
Recognition rate (%)	94.6	95.8	96.6	97.6	96.0	95.2

Furthermore, the geometric feature fusion method proposed in this paper is validated for its effectiveness in recognition., it was compared with several other methods,[4,6,7,12-14] as shown in Table 4, the proposed method demonstrates superior recognition rate and robustness across various expressions, achieving an overall recognition rate of 97.6%. This is because the proposed method combines distance and angle feature vectors that can represent 3D facial expressions for highly accurate recognition.

Table 4. Experimental comparison between this paper and other methods /%.

Algorithm	Feature Extraction Method	Database	NE	AN	DI	FE	HA	SA	SU	Over all
Quan et al.[12]	Global feature	BU-3DFE	92.0	91.0	88.0	93.0	85.0	92.0	90.1	90.2
Hariri et al.[13]	Deep features	Bosphorus	97.8	94.8	98.1	89.8	99.0	90.6	93.4	94.8
Fu et al.[14]	Shape model features	BU-3DFE		96.4	92.1	89.4	98.9	94.8	99.3	95.2
Li et al.[6]	Local feature	BU-3DFE		93.3	96.5	92.8	98.0	97.3	93.8	95.3
Guo et al.[4]	Local feature	Bosphorus		94.4	89.9	95.7	95.7	98.5	98.6	95.8
Zhang et al.[7]	Local feature	Bosphorus	100	95.7	94.1	96.8	97.5	97.7	97.1	97.0
Algorithm in this paper	Local feature	BU-3DFE	100	95.9	95.1	96.9	98.5	98.2	98.8	**97.6**

V. CONCLUSION

This paper proposes a geometric feature fusion-based method for 3D facial expression recognition. Firstly, distance and angle feature vectors that can better represent the changes in facial expression are extracted from key regions, thereby achieving high-precision facial expression recognition. Secondly, the K-NN algorithm is used to fuse the feature vector candidate set and the angle feature vector candidate set into a single feature vector, thereby reducing its dimensionality and enhancing the efficiency of facial expression recognition. The experimental results demonstrate that the proposed method has better recognition performance compared with other similar methods. However, when the key regions of facial expressions are occluded, the recognition rate is lower. In future research, we will focus on studying the problem of face occlusion.

ACKNOWLEDGMENT

This work was supported by the Fujian Province Young and Middle-aged Teachers Education Research Project (JAT220704).

REFERENCES

[1] H. Patil, A. Kothari,and K. Bhurchandi, 3-D face recognition: features, databases, algorithms and challenges, Artificial Intelligence Review,2015,44(3):393-441.

[2] Luo Changwei, Yu Jun, YU Lingyun, Overview of research progress on 3-D face recognition, J Tsinghua Univ(Sci& Technol),2021,61(1):77-88.

[3] W. Hariri, H. Tabia, N. Farah, A. Benouareth, and D. Declercq, 3D face recognition using covariance based descriptors, Pattern Recognition, 2016,78:1-7.

[4] Guo Mengli, Da Feipeng, Deng Xing, Gai Shaoyan, 3D face recognition based on keypoints and local feature, Journal of Zhejiang University(Engineering Science),2017,(51)3:584-589.

[5] M. Emambakhsh and A. Evans, Nasal patches and curves for expression-robust 3D face recognition, IEEE Trans. Pattern Anal. Mach. Intell,2017,39(5):995-1007.

[6] Li Ye, Wang Yinghui, Liu Jing, Hao Wen, Expression-insensitive 3D face recognition by the fusion of multiple subject-specific curves, Neurocomputing ,2018, 275: 1295-1307.

[7] Zhang Hongying, Yang Weimin, Wang Huisan, 3D face recognition combining local keypoints with isogeodesic curves, Laser Optoelectron,2020,22(57):1-8.

[8] Li Xiaoli, Ruan Qiuqi, Ming Yue, 3D facial expression recognition based on basic geometric features, Proceedings of the IEEE International Conference on signal processing proceedings, 2010: 1366-1369.

[9] Y. Ryan ,U.Uli,Application of the K-Nearest Neighbors (K-NN) Algorithmfor Classification of Heart Failure, Journal of Applied Intelligent System,2021,6(1):1-9.

[10] A.H.S,Jones ,C. Hardiyanti ,Case Based Reasoning using K-Nearest Neighbor with Euclidean Distance for Early Diagnosis of Personality Disorder,International Journal of Information System & Technology,2021,5(1),23-30.

[11] S.Vahidian, M. Morafah, WangWeijia, V.Kungurtsev, Chen Chen, M.Shah, and B. Lin.,Efficient distribution similarity identification in clustered federated learning via principal angles between client data subspaces. In Thirty-Seventh AAAI Conference on Artificial Intelligence, 2023,pp,10043-10052.

[12] Quan Wei, B. J. Matuszewski and L.-K. Shark, Statistical shape modelling for expression-invariant face analysis and recognition, Pattern Analysis and Applications, 2016, 19(3): 765-781.

[13] W. Hariri, N. Farah and D. K. Vishwakarma, Deep and shallow covariance feature quantization for 3D facial expression recognition, Journal of Latex Class Files, arXiv preprint, arXiv: 2105.05708v1,2021.

[14] Fu Yunfang, Deng Yujuan,Zhang Yuekui, Yang Zhengyan,Qi Ruili,Low rank tucker decomposition for 2D+3D facial expression recognition, Procedia Comput,2022 198,499-504.