Journal of Imaging Science and Technology® 66(4): 040411-1–040411-7, 2022.

Natural Feature Recognition of Multi Pose Face Images based on Augmented Reality

Wenda Xie, Jun Yu, Xiaobo Jiang, and Zongren Chen

Computer Engineering Technical College (Artificial Intelligence College), Guangdong Polytechnic Science and Technology, Zhuhai 519090, China

E-mail: xcactus@163.com

Abstract. *Traditional face feature recognition methods have a few problems, such as low accuracy of image feature recognition and high time consumption of image feature recognition. This paper introduces augmented reality technology to realize natural feature recognition of multi pose face images. The multi pose features of face image are obtained by augmented reality technology, the color normalization of face image is realized by gamma algorithm, and the natural features of multi pose face image are extracted by three-dimensional registration technology. The feature error reconstruction of multi pose face image is realized according to sparse representation, the face image samples are projected into the optimal discriminant vector space by using two-dimensional linear discriminant analysis algorithm, and the natural feature recognition of multi pose face image is completed according to augmented reality technology. The experimental results show that the comprehensive accuracy of facial image expression recognition is as high as 95%, and the average error rate of natural feature recognition is only 3%; The time taken for feature recognition is only 3.5s, which shows that the method used in this study has higher efficiency of natural feature recognition of face image.*
[DOI: 10.2352/J.ImagingSci.Technol.2022.66.4.040411]

1. INTRODUCTION

Pattern recognition is a hot subject in the field of artificial intelligence. Face recognition is a human identification technology subdivided by its basic theory. With the need of various intelligent application scenarios, face recognition technology is widely used in face image analysis and recognition. Its applications include Alipay face payment, access control systems, beauty camera and criminal investigation cases. Compared with other biometric recognition algorithms using fingerprint, iris, gait and other biometric information, face recognition is covert. It uses a non-contact camera or camera to collect images without too many operations by the collector. It is convenient and fast, has good scalability, and has many advantages, particularly non-contact operation [1, 2]. During prevention work of COVID-19 in recent two years, non-contact face recognition technology has shown great advantages. In a market research report in 2013, it was pointed out that among biometric products, face recognition products have the highest sales volume and have been showing an upward trend.

For more than 70 years, face recognition technology has been proposed and applied to all walks of life, resulting in many classical algorithms [3]. In the early algorithms, the features are designed manually, and then the classifier is selected for face recognition. In order to continuously optimize the algorithm, after years of research and development, the recognition algorithm is becoming more and more intelligent and automatic. As the cross field of pattern recognition and computing the achievements of face recognition are also constantly promoting the development and progress of other technologies [4]. Therefore, scholars in related fields have studied this.

Yu Wanbo and others improved the effect of face image recognition based on discrete cosine transform [5], obtained the detailed features of face image through DCT basis function matrix, and realized face image recognition according to the Fourier transform. This method can improve the effect of image bottom feature extraction, but the recognition rate is poor. Yan Lijuan et al. proposed a fast face recognition algorithm based on image gradient compensation [6], extracted the texture features of face image by using binocular stereo vision method, improved the resolution of face image by using Fisher faces fusion method, reduced the dimension of super-resolution face feature square image according to matching tracking algorithm, and finally realized fast face recognition through nearest neighbor classifier. This method can improve the efficiency of face recognition, but the accuracy is not high. Liu Mengjia et al. proposed a feature detection algorithm for fatigue driving face images based on SVM [7]. The texture features of face image are obtained through Dlib detection, the details of human eye fatigue are obtained according to the aspect ratio of eyes, and the feature fusion of face image is realized through support vector machine. This method can improve the positioning effect of feature points of face image, but the accuracy of face image feature recognition is poor.

To solve the above problems, this paper proposes a multi pose face image natural feature recognition method based on augmented reality (AR), which can effectively improve the accuracy and efficiency of face image feature recognition.

Received Jan. 17, 2022; accepted for publication Apr. 14, 2022; published online May 11, 2022. Associate Editor: Hamad Hamad Naeem.

1062-3701/2022/66(4)/040411/7/$25.00

2. NATURAL FEATURE RECOGNITION OF MULTI POSE FACE IMAGES BASED ON AUGMENTED REALITY

2.1 *Natural Feature Extraction from Multi Pose Face Images*

In the first step, gamma algorithm is used to realize color normalization. This step is mainly to reduce the influence of uneven illumination [8]. The calculation formula is as follows:

$$H(x,y) = H(x,y)^{\text{gamma}}, \tag{1}$$

where $H(x,y)$ represents the gray value of pixel points (x,y).

In the second step, the local contour information of the image is obtained by calculating the pixel gradient [9]. The gradient is calculated as follows:

$$G_x(x,y) = H(x+1,y) - H(x-1,y), \tag{2}$$

$$G_y(x,y) = H(x,y+1) - H(x,y-1), \tag{3}$$

where $G_x(x,y)$ represents the gradient value in the x direction at the pixel point (x,y), and $G_y(x,y)$ represents the gradient value in the y direction at the pixel point (x,y) [10].

This is followed by the calculation of the gradient amplitude and gradient direction values of pixel points (x,y). The gradient amplitude is $G(x,y)$, and the gradient direction value is $\alpha(x,y)$. The calculation formula is as follows:

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{4}$$

$$\alpha(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right), \tag{5}$$

The third step is to calculate the gradient direction histogram in the cell unit. This step is performed primarily to reduce the sensitivity of feature extraction to face shape [11].

The fourth step is to normalize the gradient direction histogram in the block in the block unit. The specific operation is to organize adjacent $t*t$ cells into a block, connect the gradient direction histograms in the cells in series to obtain the gradient direction histogram of the block, and then normalize the block [12–15].

The fifth step is to obtain the natural features of the whole multi pose face image. The gradient direction histograms of all blocks are connected in series to obtain the natural features of the whole image.

3. SPARSE REPRESENTATION OF MULTI POSE FACE IMAGE FEATURES

Suppose that the matrix $\boldsymbol{A}_j$ represents all training samples of class j, with $A = [A_1, A_2, \ldots, A_M]$, A represents the matrix composed of all training samples, and the number of sample categories is M. The test sample y can be linearly expressed as $Aw = y$. The sparse solution can be obtained from formula (6) [16].

$$w_1^* = \arg\min \|w\|_1 \quad \text{s.t.} \quad y = Aw. \tag{6}$$

In the formula, $\|\cdot\|_1$ represents $L1$ norm. Although it has good sparsity by solving $L1$ norm, it has high algorithm complexity. Therefore, $L2$ norm is used for sparse solution, which is also called collaborative representation classification method, as shown in formula (7) [17].

$$w_2^* = \arg\min \|w\|_2 \quad \text{s.t.} \quad y = Aw. \tag{7}$$

The reconstructed prediction sample $y_j^* = A\delta_j(w_2^*)$ is obtained by sparse solution and all training samples of each class $\delta J(w*2)$, y_j^* represents the test sample after class j reconstruction [18]. The reconstruction error of class j corresponding to it can be expressed as shown in formula (8).

$$R_j(y) = \left\|y - y_j^*\right\|_2 = \left\|y - A\delta_j(w_2^*)\right\|_2. \tag{8}$$

Another more specific expression of reconstruction error is expressed by the formula (9):

$$R_j(y) = \left\| y - \sum_{j=1}^{c} \hat{a}\,(j-1)_{n+i}\, x_j^i \right\|_2, \tag{9}$$

where $R_j(y)$ can be regarded as a distance measure, indicating the similarity between the test sample y and the $j(1 \le j \le M)$ training sample. $\hat{a}(j-1)_{n+i}$ represents the $(j-1)_{n+i}$ element of A_j.

3.1 *Two Dimensional Linear Discriminant Analysis of Training Samples based on Augmented Reality*

Augmented reality technology can recognize the position, direction and size of markers in the image, and overlay the three-dimensional virtual image to the corresponding position of the real image. Intelligent display technology presents the results of the integration of virtual scene and real scene in front of users; and real time interaction technology realizes the interaction between users and augmented reality system through human-computer interaction technology. Firstly, for multi pose face image acquisition in this study, the image is collected by the panoramic annular imaging system, and the original circular compressed image is expanded according to the camera model and internal parameters. The panoramic annular imaging system can obtain a 360° field of view, and the collected images can be used to realize augmented reality and expression recognition. Then, feature interactive processing is realized by augmented reality. The mark image is recognized by 3D registration technology based on natural features, and the 3D shape is projected onto the image space. The system runs on the notebook computer, the fusion image is projected to the display screen, and the model control is realized through the keyboard and mouse, including the rotation and scaling of the three-dimensional model. Finally, the natural features of multi pose face images are recognized. The multi pose face image natural feature recognition module first carries out face detection, and then carries out expression recognition on the area where the face is located to judge the user's state. The recognition results are interactive through sound.

Using the two-dimensional linear discriminant analysis algorithm, the image samples are projected into the optimal discriminant vector space, that is, the non separable image data samples in the original space are projected into a linear separable space. It can reduce the dimension of the original data. Moreover, after projection in linear separable space, the dispersion distance between different classes is larger.

Suppose that the N samples of the data set come from M categories respectively, S_b represents the inter class dispersion matrix, and the data set is $X_j \in R^{m\times n}(i = 1, 2, \ldots, N)$, where $m \times n$ represents m rows and n columns of an image in the data set, and M_j of class J in the data set has r_j samples, and S_w represents the intra class dispersion matrix.

The purpose of 2DLDA is to find the projection matrix W [19]. The objective function of multi pose face image projection matrix of DLDA algorithm is shown in formula (10).

$$\begin{aligned} W_{opt} &= \arg \max_{W^T W=I} \frac{\sum_{j=1}^{M} r_j \left\| W^T \overline{X}_j - \overline{X} \right\|_F^2}{\sum_{j=1}^{M} \sum_{i\in c_i} r_j \left\| W^T X_i - \overline{X}_j \right\|_F^2} \\ &= \arg \max_{W^T W=I} \frac{Tr\left(W^T S_b W\right)}{Tr\left(W^T S_w W\right)}. \end{aligned} \tag{10}$$

3.2 *Natural Feature Recognition of Multi Pose Face Images*

The three-dimensional object in OpenGL consists of six faces, which are left faces, $x = 1$; Right plane, $x = r$; Bottom = face, $y = b$; Top plane, $y = t$; Near surface: $z = n$; Far face: $z = f$. In OpenGL, the projection matrix formula is as follows:

$$P = \begin{bmatrix} \frac{2f_x}{w} & 0 & 1 - \frac{-2c_x}{w} & 0 \\ 0 & \frac{2f_y}{h} & \frac{2c_y}{w} - 1 & 0 \\ 0 & 0 & -\frac{f+n}{f-n} & -\frac{2f}{f-n} \\ 0 & 0 & -1 & 0 \end{bmatrix}. \tag{11}$$

where, f_x and f_y are the focal lengths of the camera in the x-axis and y-axis directions, w is the width of the image, h is the height of the image, and f and n represent the near and far plane positions of the z-axis. Projection transformation can display 3D models in OpenGL system. The 3D model in OpenGL uses obj format files.

Obj file format is a simple file format for representing 3D geometry. The 3D model is represented by vertex coordinates, UV coordinates of each vertex texture, vertex normal vector, vertex coordinates of polygons and texture UV coordinate sequence. Use the GLEW library to realize the OpenGL function. GLEW is a cross platform C++ extension library based on OpenGL graphics interface, which can automatically identify all OpenGL advanced extension functions supported by the current platform. The movement, rotation and scaling of images in OpenGL are realized by matrix multiplication, so various operations on graphics are essentially matrix coordinate transformation. However, OpenGL does not have its own matrix and vector calculation library, so GLM (OpenGL math) library is introduced. GLM is an OpenGL mathematics library, which can quickly realize various mathematical matrix operations. OpenGL provides functions to control the model. The calling function can monitor the operation of keyboard and mouse, and then realize the control of rotation, scaling and translation of the model. In this study, glfwgetkey() is used to obtain external device input, translate() is used to translate, and scale() is used to zoom.

For the input, original low resolution image, m groups of deep face features are obtained by using the face feature construction method, and then fused by using the feature fusion factor. The fusion features are obtained by using the feature fusion function. The trained feature fusion w and error b are applied to the depth feature to calculate the fusion feature.

$$\phi(f) = w \times f + b. \tag{12}$$

Among them, function ϕ represents a feature fusion function for training, f is a group of depth features extracted by ResNet-50 network, w is the weight of feature fusion factor, and b is the error of feature fusion factor.

This is followed by addition of m groups of fused features to obtain the deeply fused face features.

$$F = \left(\phi(f_1) + \phi(f_2) + \ldots + \phi(f_m)\right). \tag{13}$$

The function ϕ represents the feature fusion function designed in this study for training, F is the final face super score feature, and $f_1, f_2, \cdots, f_m$ represents one group of depth super score face features and $m - 1$ group of depth fractional order face features respectively.

When using the feature fusion factor for feature fusion, we impose constraints on the weight of the feature fusion factor. The constraints are shown in Eq. (14), so that the dimensions and norms of the fused features remain unchanged:

$$w_1 + w_2 + \cdots + w_m = 1. \tag{14}$$

At the same time, this study also sets the error $b_1, b_2, \ldots, \mathrm{b}_m$ of the fusion factor to 0 to prevent unnecessary interference to feature fusion. After obtaining the depth feature rich in face features, the feature can be used for face recognition.

Using *ResNet*-50 as the feature extraction network, m sets of face image feature vectors are obtained respectively, and these features are fused. When training the algorithm proposed in this study, triple loss is selected as the loss function for training.

$$\begin{aligned} f_{trip\_+loss} = \sum_{i=1}^{n} \max \Big[&\|F(x_i^a) - F(x_i^p)\|_2^2 - \|F(x_i^a) \\ &- F(x_i^n)\|_2^2 + a, 0 \Big], \end{aligned} \tag{15}$$

where, n is the number of training samples, $F(\cdot)$ is the algorithm model proposed in this paper, x_i^a represents the current training samples, x_i^p represents the face samples

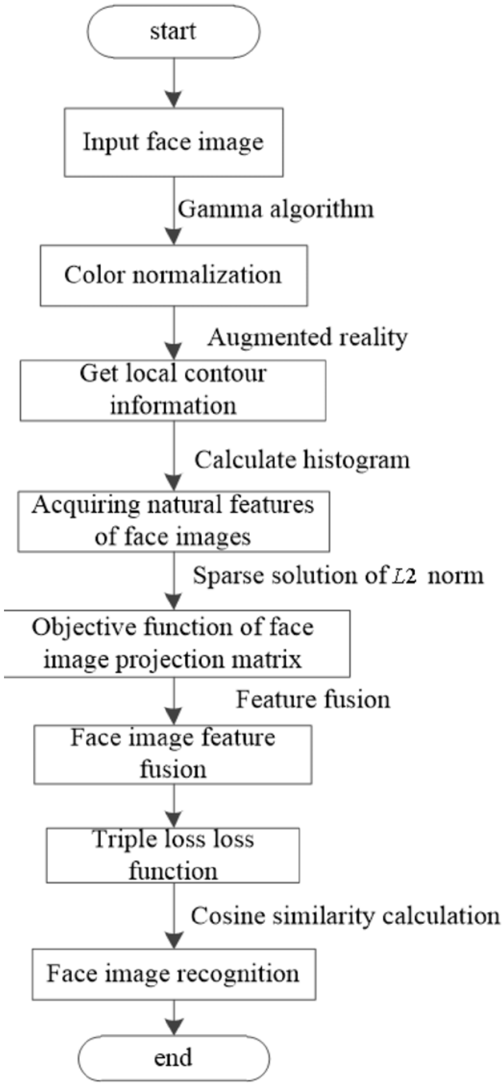

Figure 1. Face image recognition algorithm.

similar to the current samples, x_i^n represents the face samples of different classes from the current samples, $\|\cdot\|$ is the Euclidean distance calculation function, and a represents the distance interval.

In the test, the face image is input into the trained network model, and the cosine similarity is calculated with the reference image vector in the face image database to complete the recognition. The flow chart of image recognition algorithm in this study is shown in Figure 1.

As shown in Fig. 1, this study first inputs the face image, uses the gamma algorithm to normalize the color of the face image, obtains the local contour information of the image through augmented reality, calculates the gradient direction histogram to obtain the natural features of the multi pose face image and the the objective function of the projection matrix of the face image through the thinning solution of $L2$ norm. The face image is fused by feature fusion factor. Finally, the face image training samples are trained through the triple loss function, and the cosine similarity is calculated to complete the multi pose face image recognition.

Table I. Expression classification mapping.

Primitive expression classification	New expression classification
Happy	Positive emotional expression
Surprised	
Angry	Negative emotional expression
Hate	
Fear	
Sad	
Neutral	Neutral emotional expression

4. EXPERIMENT

4.1 *Data Acquisition*

Facial expressions can be divided into seven types: anger, disgust, fear, happiness, sadness, surprise and neutrality. Considering the application scenario, the expression is further simplified into positive emotional expression, negative emotional expression and neutral emotional expression. As shown in Table I, positive emotional expressions include happiness and surprise, negative emotional expressions include anger, disgust, fear and sadness, and neutral emotional expressions include neutrality.

Collect 100 panoramic photos of positive emotional expression, 100 panoramic photos of negative emotional expression and 100 panoramic photos of neutral emotional expression. Expand the original image and intercept the image of the area where the face is located to form image data, and mark manually according to the image expression. The image acquisition of the whole database is divided into two stages. A certain time interval is required between the two stages, and 13 photos are collected for each person in each stage. Among the 13 photos, 7 were taken under different lighting and facial expressions, and 6 were taken under partial occlusion with glasses, sunglasses, etc. The images of some AR data sets are shown in Figure 2. All occluded objects in the figure are regarded as noise in the recognition process, so it is not necessary to add noise manually. The size of each picture is cut from the original $576 * 768$ to $50 * 40$ used in the experiment. Finally, the facial expressions are scored according to the expression classification method shown in Table I.

4.2 *Result Analysis*

4.2.1 *Comprehensive accuracy of facial image expression recognition*

The classification of Jaffe data sets is realized by AR technology. The experimental results are shown in Table II.

In order to verify the effectiveness of this method, literature [5], literature [6], literature [7] and this method are used to detect the comprehensive accuracy of facial image expression recognition. The detection results are shown in Figure 3.

Figure 2. Schematic diagram of face image data.

Table II. Expression recognition accuracy of face image (%).

Expression classification	Accuracy
Positive emotional expression	96.6
Neutral emotional expression	95.0
Negative emotional expression	97.2

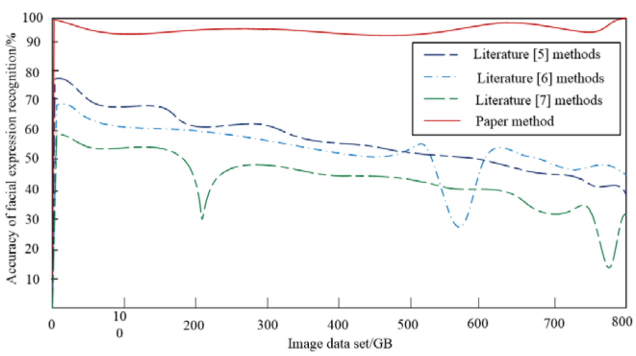

Figure 3. Comprehensive accuracy of facial image expression recognition

According to the analysis of Fig. 3, when the image data set is 100GB, the comprehensive accuracy of facial image expression recognition of literature [5] method is 68%, the comprehensive accuracy of facial image expression recognition of literature [6] method is 62%, the comprehensive accuracy of facial image expression recognition of literature [7] method is 55%, and the comprehensive accuracy of facial image expression recognition of the proposed method is 95%; When the image data set is 500GB, the comprehensive accuracy of facial image expression recognition of literature [5] method is 53%, the comprehensive accuracy of facial image expression recognition of literature [6] method is

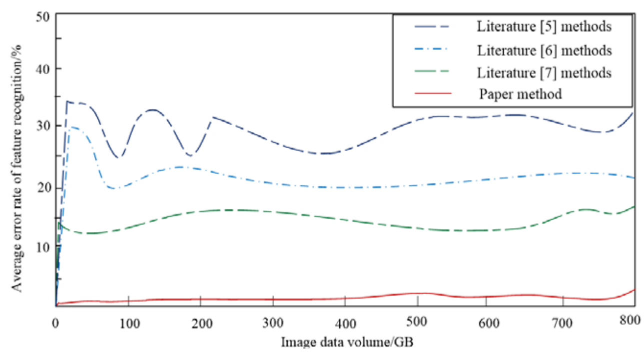

Figure 4. Average error rate of natural feature recognition of face image

58%, the comprehensive accuracy of facial image expression recognition of literature [7] method is 42%, and the comprehensive accuracy of facial image expression recognition of the proposed method is 95%. The comprehensive accuracy of face image expression recognition of the method proposed in this study is much higher than that of other methods and has a better recognition effect. This is because the study method uses the Gamma algorithm to normalize the color of the face image, and realizes the feature error reconstruction of the multi-pose face image according to the sparse representation, which effectively improves the expression recognition effect of the face image.

4.2.2 *Average error rate of image feature recognition*

In order to verify the effectiveness of this method, literature [5], literature [6], literature [7] and the proposed method are used to recognize the average error rate of natural feature of face image, and the detection results are shown in Figure 4.

Based on analysis of Fig. 4, when the amount of face image data is 100GB, the average error rate of natural feature recognition of the method in literature [5] is 31%, the average

Table III. Natural feature recognition time of face image.

throughput/kbit	Time for natural feature recognition of face images/s			
	Literature [5] methods	Literature [6] methods	Literature [7] methods	Study method
1000	28.43	36.28	38.52	0.2
2000	38.87	45.29	42.76	1.7
3000	42.69	55.86	57.43	2.1
4000	48.28	69.42	66.19	2.8
5000	63.89	76.22	73.87	3.2
6000	78.95	88.68	79.43	3.5

error rate of natural feature recognition of the method in literature [6] is 22%, the average error rate of natural feature recognition of the method in literature [7] is 13%, and the average error rate of natural feature recognition of the method in this study is only 1%; When the amount of face image data is 500GB, the average error rate of natural feature recognition of literature [5] method is 34%, the average error rate of natural feature recognition of literature [6] method is 21%, the average error rate of natural feature recognition of literature [7] method is 13%, and the average error rate of natural feature recognition of this method is 3%; The average error rate of the method in this study is significantly lower, because this study uses the three-dimensional registration technology to extract the natural features of the multi-pose face image; uses the two-dimensional linear discriminant analysis algorithm to project the face image samples into the optimal discriminant vector space, according to the enhanced Realistic technology completes the recognition of natural features of multi-pose face images, which makes the method in this study have a better recognition effect of natural features of face images.

4.2.3 *Time for Natural Feature Recognition of Face Images*

In order to verify the effectiveness of this method, when using the methods in literature [5], literature [6], literature [7] and the proposed method for natural feature recognition of face image, the detection results are shown in Table III.

Sort out the data in Table III to get Figure 5.

According to the analysis of Table III and Fig. 5, When the throughput is 1000kbit, the time for natural feature recognition of face image in the method of literature [5] is 28.43 s, the time for natural feature recognition of face image in the method of literature [6] is 36.28 s, the time for natural feature recognition of face image in the method of literature [7] is 38.52 s, and the time for natural feature recognition of face image in the method of this study is 0.2S; When the throughput is 6000 kbit, the time for natural feature recognition of face image in literature [5] is 78.95 s, the time for natural feature recognition of face image in literature [6] is 88.68 s, the time for natural feature recognition of face image in literature [7] is 79.43 s, and the time for natural feature recognition of face image in the study method is 3.5 s. The method in this study has a higher recognition efficiency of

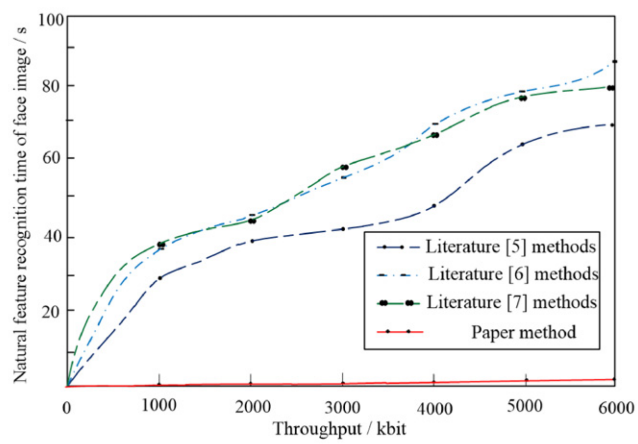

Figure 5. Natural feature recognition time of face image.

natural features of face images. This is because this study uses augmented reality technology to obtain multi-pose features of face images, and uses three-dimensional registration technology to extract natural features of multi-pose face images. This makes this method highly efficient in face image natural feature recognition.

5. CONCLUSION

This paper presents a natural feature recognition method of multi pose face image based on augmented reality. The color normalization of face image is realized by gamma algorithm, and the natural features of multi pose face image are extracted by three-dimensional registration technology; The feature error reconstruction of multi pose face image is realized according to sparse representation, and the natural feature recognition of multi pose face image is completed according to augmented reality technology. The following conclusions are obtained through experiments:

(1) When the image data set is 500GB, the comprehensive accuracy of facial image expression recognition of this method is 95%, which shows that this method has a good recognition effect.
(2) When the amount of face image data is 500GB, the average error rate of natural feature recognition is 3%; the average error rate of this method is obviously low

indicating that this method has good natural feature recognition effect of face image.

(3) When the throughput is 6000kbit, and the time of face image natural feature recognition based on this method is 3.5s. This method has higher efficiency of natural feature recognition of face image.

ACKNOWLEDGMENTS

This work was supported in part by the Key scientific research platforms and projects of colleges and universities of Guangdong Provincial Department of Education under Grant no. 2021ZDZX1147.

Data Availability

The data that support the findings of this study are available upon request from the corresponding author.

REFERENCES

[1] C. Yubin, "Research on similar expression recognition of video moving face images based on improved kernel discrimination algorithm," J. Qiqihar Univ. (Natural Science Edn.) **37**, 36-40 + 46 (2021).

[2] Z. Zhen, L. Yunong, R. Shuai, and Z. Degang, " Differential histogram shift Lossless Information Hiding Algorithm Based on LBP face texture feature," Comput. Appl. Res. **37**, 1774–1778 (2020).

[3] Z. Yingying, G. mingcen, Y. Lingwei, Z. Beixin, and H. Jinfeng, "Fuzzy face image feature similarity recognition based on ternary space fusion," Comput. Age 66–69 (2021).

[4] W. Hongxing, H. yongyang, and D. Chao, "Research and implementation of face recognition algorithm based on LBP and elm," J. Henan Univ. Technol. (Natural Science Edition) **40**, 139–145 (2021).

[5] Y. Wanbo, W. Xiangxiang, and W. Daqing, "Face image recognition based on discrete cosine transform basis function iteration," J. Graph. **41**, 88–92 (2020).

[6] Y. Lijuan and Z. Yanhu, "Fast face recognition algorithm based on image gradient compensation," Comput. Syst. Appl. **29**, 194–201 (2020).

[7] L. Mengjia and Z. Jianguo, "Feature detection algorithm of fatigue driving face image based on SVM," Semicond. Optoelectron. **41**, 278–282 (2020).

[8] W. Jing and Y. Yuan, "Implementation of CNN face recognition method based on feature distance," J. Xinjiang Norm. Univ. (Natural Science Edition) **40**, 67–70 (2021).

[9] A. Xiaoning, W. Zhiwen, Z. Canlong, G. Jiaying, and L. Qiuling, "Face feature annotation and recognition based on Hidden Markov model," J. Guangxi Univ. Sci. Technol. **31**, 118–125 (2020).

[10] L. Zepeng and Y. Tinglai, "Facial expression feature recognition based on local binary pattern and limit learning machine classification algorithm," Electron. World 78–79 (2020).

[11] X. Yiming, Y. Yingying, C. Liying, P. Jinpeng, L. Yue, and X. Xiangyu, "Research on face detection algorithm based on Haar AdaBoost," Equip. Manuf. Technol. 67-70–+ 75 (2020).

[12] Z. Huaping, Z. Yiyi, S. Kele, Q. Huangli, and G. Haixia, "Facial expression feature extraction method based on ENM Gabor differential weight," Comput. Appl. Softw. **37**, 184-189 + 212 (2020).

[13] Q. Ma and Y. Wang, "Application of embedded system and artificial intelligence platform in Taekwondo image feature recognition," J. Ambient Intell. Humanized Comput. **18**, 52–69 (2021).

[14] S. Zou, H. Chen, H. Zhou, and J. Chen, "An intelligent image feature recognition algorithm with hierarchical attribute constraints based on weak supervision and label correlation," IEEE Access **8**, 105744–105753 (2020).

[15] K. Qian, L. Tian, Y. Liu, X. Wen, and J. Bao, "Image robust recognition based on feature-entropy-oriented differential fusion capsule network," Appl. Intell. **18**, 1–10 (2020).

[16] M. Jingyi, T. Zhang, J. Guodong, Y. Wenjun, and Y. Bin, "Ground-based cloud image recognition system based on multi-CNN and feature screening and fusion," IEEE Access **8**, 1–19 (2020).

[17] Y. Sun, Y. Weng, B. Luo, G. Li, B. Tao, D. Jiang, and D. Chen, "Gesture recognition algorithm based on multi-scale feature fusion in RGB-D images," IET Image Process. **18**, 212–226 (2020).

[18] B. Y. Panchal, G. Chauhan, S. R. Panchal, and U. M. Chaudhari, "An investigation on feature and text extraction from images using image recognition in Android," Mater. Today: Proc. **51**, 1–11 (2022).

[19] E. Asali, F. Shenavarmasouleh, F. G. Mohammadi, P. S. Suresh, and H. R. Arabnia, "DeepMSRF: A novel deep multimodal speaker recognition framework with feature selection," in *Advances in Computer Vision and Computational Biology. Transactions on Computational Science and Computational Intelligence*, edited by H. R. Arabnia, L. Deligiannidis, H. Shouno, F. G. Tinetti, and Q. N. Tran (Springer, Cham, 2021).