3D Face Recognition Method Based on Deep Convolutional Neural Network

Jianying Feng, Qian Guo, Yudong Guan, Mengdie Wu, Xingrui Zhang and Chunli Ti

Abstract In 2D face recognition, result may suffer from the impact of varying pose, expression, and illumination conditions. However, 3D face recognition utilizes depth information to enhance systematic robustness. Thus, an improved deep convolutional neural network (DCNN) combined with softmax classifier to identify face is trained. First, the preprocessing of color image and depth map is different in removing redundant information. Then, the feature extraction networks for 2D face image and depth map are, respectively, build with the principle of recognition rate maximization, and parameters about neural networks reset by a series of tests, in order to acquire higher recognition rate. At last, the fusion of two feature layers is the final input of artificial neural network (ANN) recognition system, which is followed by a 64-way softmax output. Experimental results demonstrate that it is effective in improving recognition rate.

Keywords 3D face recognition · Depth map · Deep convolutional neural network Feature extraction · Feature fusion

---

J. Feng (✉) · Y. Guan · X. Zhang · C. Ti
School of Electronics and Information Engineering, Harbin Institute of Technology,
Xidazhi Street. 92, Harbin 150001, China
e-mail: fjy_fengjianying@163.com

Y. Guan
e-mail: 157673503@qq.com

X. Zhang
e-mail: 16S105149@stu.hit.edu.cn

C. Ti
e-mail: 823231451@qq.com

Q. Guo
Beijing Automation Control Equipment Research Institute, Beijing 100074, China
e-mail: 13810106152@139.com

M. Wu
School of Information Engineering Branch, Yangling Vocational & Technical College,
Weihuilu. 24, Yangling 712100, Shanxi, China
e-mail: 13131724330@163.com

B. K. Panigrahi et al. (eds.), *Smart Innovations in Communication and Computational Sciences*, Advances in Intelligent Systems and Computing 670,
https://doi.org/10.1007/978-981-10-8971-8_12

1 Introduction

Face recognition has been widely applied because of its characteristics of being easy to acquire, direct, interaction, and so on [1]. It is a valuable research direction for pattern recognition and machine learning. Two-dimensional face recognition is relatively complete at present, but it is not enough robust when it confronts various conditions. Three-dimensional face recognition utilizes depth information to avoid the inflation of environmental factors, which exist in 2D face recognition. More and more researchers focus on 3D face recognition, while the recognition rate of traditional superficial layer methods has trend saturation, especially when the training sets are enormous. Thus, some people try to utilize neural network to solve the above-mentioned problem [2, 3]. It is proposed that the cascade correlation neural network with multi-core programming model realizes 3D facial recognition system [4]. A method for automatically recognizing expression using DCNN is proposed. This method can reduce the time of extracting feature by general-purpose graphic processing unit [5]. Traditional face recognition methods are susceptible to various illuminations, different expressions, and other changeable conditions. Multilayer constructions, especially deep convolutional network, can have flexible reaction to these nonlinearity problems.

In this paper, DCNN is applied in 3D face recognition. Firstly, appropriate methods are taken to preprocess original data, including elimination of redundancy information and normalization. Secondly, an improved DCNN combined with softmax classifier is trained. As the principle of recognition rate maximization, the processions of extracting 2D and 3D features are independent. Moreover, the output of two feature layers is considered as the input of ANN recognition system. Finally, the recognition system is tested and compared with other methods. Excellent experiment performances on testing set show the great superiority of this method.

The rest of the paper is composed as follows: The second part elaborates the construction of DCNN in detail; the corresponding experiments are demonstrated in third part; and conclusion will be summarized in the last part.

2 3D Face Recognition Method Based on DCNN

2.1 Construction of DCNN

DCNN for color image is shown in Fig. 1, where the input image is 64×64, the first convolutional layer with 15 kernels is 5×5, the second convolutional layer includes 30 kernels of size 4×4, the size of the third convolutional layer with 45 kernels also is 4×4, and the kernel of two max-pooling layers is defined as 2×2, 3×3, respectively. Figure 2 displays the DCNN of exacting depth feature. The kernels' size of two convolutional layers both is 5×5, and two max-pooling sampling layers is 2×2 Considering the speed and complexity computation, the activation function of two DCNNs is ReLU rather than Sigmoid [6].

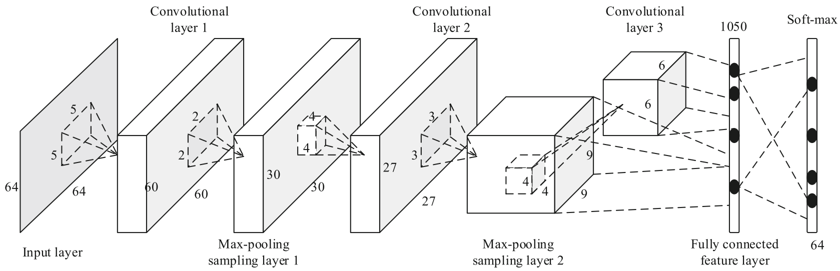


Fig. 1 Construction of DCNN for color image

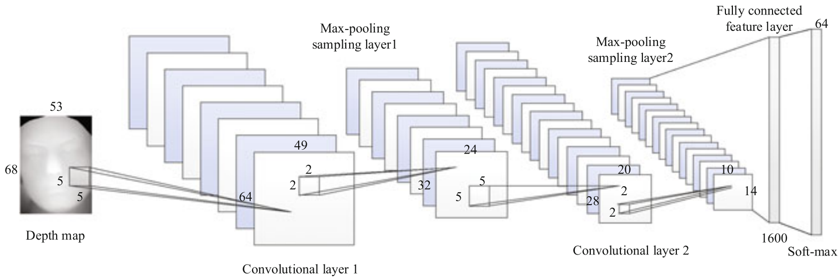


Fig. 2 Construction of DCNN for depth map

2.2 Softmax Regression Layer and Training Algorithm

Softmax layer outputs n parameters, which express the probability of n kinds of face samples. Given the predicted label $y \in \{1, 2, \ldots, n\}$, the probability of object belonging to the ith classify is

$$\phi_i = P(y = i|\phi) \tag{1}$$

where $\sum_{i=1}^{n} \phi_i = 1$

The predicted value of nth classify is

$$P(y = n|\phi) = \sum_{i=1}^{n1} \phi_i \tag{2}$$

Thus,

$$\phi_i = \frac{\exp\,(\theta_i^T x)}{\sum_{j=1}^{n} \exp\,(\theta_i^T x)} \quad i \in \{1, 2, \ldots, n\} \tag{3}$$

where $\theta_j = \{W_j, b_j\}, j \in \{1, 2, \ldots, n\}$ are the parameters of softmax.

The cost function of this network based on maximum likelihood estimation is

$$\begin{aligned} J(\theta) &= \frac{1}{K}\sum_{i=1}^{K} \log P(y_i \big| x_i^{(5)} \theta) \\ &= \frac{1}{K}\sum_{i=1}^{K} \log \prod_{j=1}^{n} \left(\frac{\exp\left(\theta_j^T x_i^{(5)}\right)}{\sum_j^n \exp\left(\theta_j^T x_i^{(5)}\right)} \right)^{\text{isture}(y_i = j)} \\ &= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n} \text{isture}(y_i = j) \cdot \log \frac{\exp\left(\theta_j^T x_i^{(5)}\right)}{\sum_j^n \exp\left(\theta_j^T x_i^{(5)}\right)} \end{aligned} \tag{4}$$

where θ is a alterable vector, K is the total number of samples, $x_i^{(5)}$ is the ith input, $i \in \{1, 2, \ldots, m\}$, y_i is the corresponding predicted label. The final training target is getting the minimum of cost function. The target function is defined as

$$\arg \min_{\theta} \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n} \text{isture}(y_i = j) \cdot \log \frac{\exp\left(\theta_j^T x_i^{(5)}\right)}{\sum_j^n \exp\left(\theta_j^T x_i^{(5)}\right)}, \quad i \in (1, 2, \ldots, n) \tag{5}$$

The partial derivation of Eq. (5) is

$$\nabla_{\theta_i} J(\theta) = \frac{1}{K}\sum_{i=1}^{K} \left[x_i^{(5)} \cdot \left(\text{isture}(y_i = j) P\left(y_i = j \big| x_i^{(5)} \theta \right) \right) \right] \tag{6}$$

Softmax is n dimension, but its degree of freedom is $n - 1$, and thus, there is redundancy δ. Redundancy leads possibly to overlarge update. Usually, λ is added to cost function. Give an example as follows

$$J(\theta) = \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n} \left[\text{isture}(y_i = j) \cdot \log \frac{\exp\left(\theta_j^T x_i^{(5)}\right)}{\sum_j^n \exp\left(\theta_j^T x_i^{(5)}\right)} \right] \frac{\lambda}{2} \sum_{i=1}^{n}\sum_{j=0}^{120} \theta_{ij}^2 \tag{7}$$

The partial derivation of Eq. (7) is

$$\nabla_{\theta_i} J(\theta) = \frac{1}{K}\sum_{i=1}^{K} \left[x_i^{(5)} \cdot \left(\text{isture}(y_i = j) P\left(y_i = j \big| x_i^{(5)} \theta \right) \right) \right] \lambda \cdot \theta_j \tag{8}$$

It can be proved that $J(\theta)$ is a convex function. In order to reduce computational expense and speed up the training procession, stochastic gradient descent algorithm is taken. Training samples are randomly divided into specific groups to reconstitute the cost function with these groups.

In the real training, there is a coefficient of momentum to control iteration,

$$v^{(i1)} = 09v^{(i)}\alpha\nabla_{\theta_i}J(\theta)^{(i)} \tag{9}$$

where 0.9 is a constant, v is a momentum, α is studying rate.

2.3 Feature Fusion

The fusion of two feature layers is the final input of ANN recognition system. The structure is summarized in Fig. 3, where the input of 2D face feature is 1050 dimension, depth feature is 1600 dimension, partially connected layers is the half of corresponding input, and the final output layer is a 64-way softmax. Moreover, the number of fully connected layer is set as 450 by tests.

3 Experimental Results and Discussion

The experiment has been conducted using CASIA-3D FaceV1 collected by the Chinese Academy of Sciences' Institute of Automation (CASIA) [7]. The database contains 4624 scans of 123 people; each person has 37 or 38 scans, where the one who wears glasses has one additional scan. Furthermore, it considers various poses, illuminations, and expressions. Three-dimensional facial data is saved as wrl file,

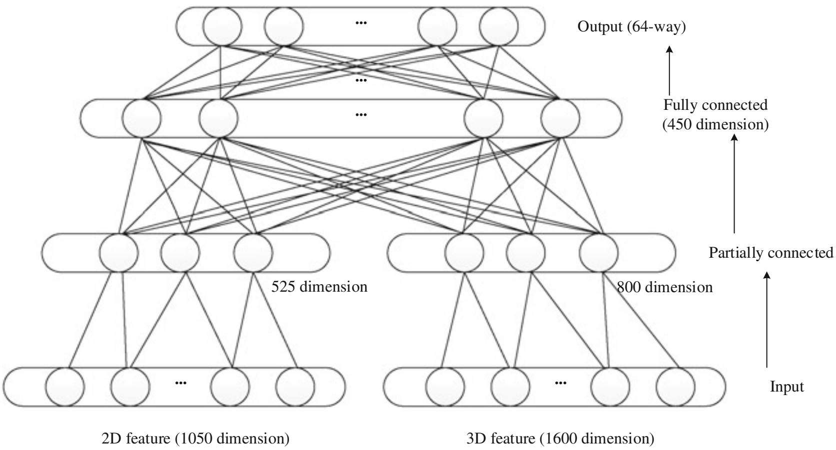


Fig. 3 Structure of ANN

Fig. 4 Some depth maps after preprocessing

which gives some information such as the number and space coordinates of point cloud, each point color (RGB) and normal vector, and the number of triangular patches. There are 64 persons selected as the final experimental data, and everyone includes 16 color images and 10 depth maps.

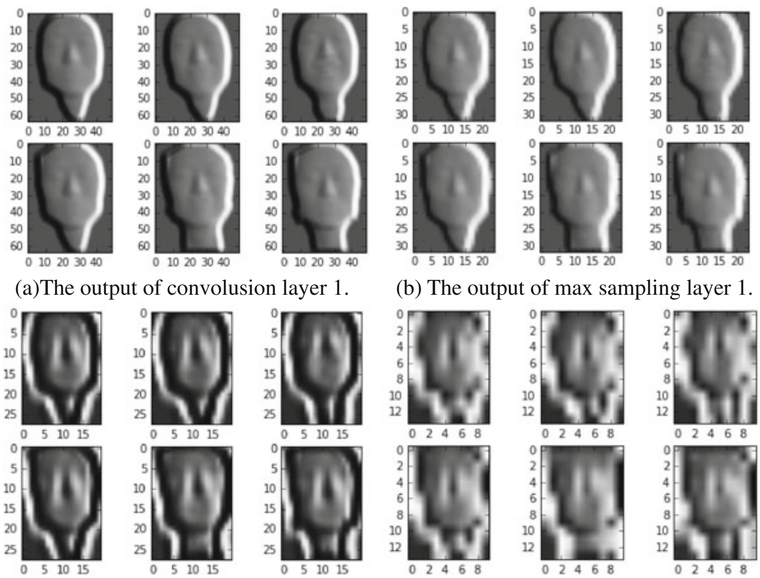
AdaBoost algorithm combined with Haar feature is used to remove redundancy information of 2D face images [8, 9]. Convert color images to gray images, then normalized 64×64. The 3D grid model is built according to the point cloud data to get depth map. Then, Otsu's method is applied to remove redundancy [10]. Figure 4 shows some depth maps which are normalized to 68×53.

Learning rate can be regarded as the range of parameters update. Data preprocessed is divided into training set, verifying set, and testing set. To find the most suitable learning rate, respectively, checkout learning rate $\alpha = 01, 007, 006 \cdot 005, 004, 001$ Finally, learning rate is set to 0.06 according to the maximization of recognition rate.

Different features' number between DCNN and softmax can result in different recognition rate. So based on a series of test, 2D and 3D feature numbers are, respectively, reset to 1050 and 1600, namely the dimensions of the input layer. In addition, Fig. 5 displays 3D feature map of every layer output for better comprehension. Two-dimensional feature map is not given to economize space, and it also makes 3D feature more and more abstract with the increasing of layers.

To illustrate the stability of the method, five different test sets were used for testing. The recognition rate is 96.87, 98.44, 100.00, 95.31, and 98.44%. There are two reasons for the different recognition rate. Objective reason is that pose and illumination conditions are different. Subjective reason is that the preprocessing is not enough ideal.

In the end, this method is compared with single 2D recognition, depth recognition, and some others. The final recognition rate with CASIA database is displayed in Table 1. From the data can be seen that this method is better than other methods when system is based on 2D feature or 3D feature. The recognition rate with fusion feature is higher about 5% than 2D feature in this method. Experimental results are commendable.



(a)The output of convolusion layer 1. (b) The output of max sampling layer 1.

(c) The output of convolusion layer 2. (d) The output of max sampling layer 2.

Fig. 5 Visualization output of depth feature

Table 1 Recognition rate of different methods

	Recognition rate (%) (2D)	Recognition rate (%) (3D)	Recognition rate (%) (2D and 3D)
PCA	80.100	70.430	–
ICA&SVM	87.000	78.940	–
Traditional CNN	83.570	78.280	–
Improved Fisher	91.800	80.430	–
Method in this paper	93.750	85.938	98.440

4 Conclusion

Face recognition has been widely used in many fields, but 2D recognition technology is confronted with the bottleneck of development and extension as it is instable of constantly changing factors. Facial depth map embodies 3D feature, which is beneficial for face recognition. In this paper, a 3D face recognition method with DCNN is designed. The fusion of 2D and 3D feature as the final input of softmax makes the best of all information. Experimental results demonstrate that it is effective in improving the recognition rate. With the development of 3D technology, 3D face recognition will be popularized in human–computer interaction, public security, entertainment, and so on.

References

1. Holland, C.D., Komogortsev, O.V.: Complex Eye Movement Pattern Biometrics: The Effects of Environment and Stimulus. Transactions on Information Forensics & Security. 8(12), 2115–2126 (2013).
2. Jadon, S., Kumar, M., Rathi, Y.: Face recognition using Som Neural Network with Ddct facial feature extraction techniques. In: 2nd IEEE International Conference on Communications and Signal Processing, pp. 1070–1074. IEEE Press, Chennai (2015).
3. Boughrara, H., Chtourou, M., Amar, C.B., et. al.: Facial expression recognition based on a mlp neural network using constructive training algorithm. Multimedia Tools and Applications. 75, 1–23 (2016).
4. Alqatawneh, S.M., Jaber, K.M.: Parallel Cascade Correlation Neural Network Methods for 3D Facial Recognition: A Preliminary Study. Journal of Computer & Communications. 3, 354–373 (2015).
5. Mayya, V., Pai, R.M., Pai, M.M.M.: Automatic Facial Expression Recognition Using DCNN. Procedia Computer Science. 93, 453–461 (2016).
6. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Huang, T.W., Zeng, Z.G., Li, C.D. (eds.) Euro-Par 2012. NIP, vol. 25, pp. 1097–1105. Springer, Heidelberg (2012).
7. Chinese Academy of Sciences' Institute of Automation, http://biometrics.idealtest.org.
8. Erdem C.E., Ulukaya S., Karaali A., et al. Combining Haar Feature and skin color based classifiers for face detection. In: 36th IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1497–1500. IEEE Press, Prague (2011).
9. Huang, C.C., Tsai, C.Y., Yang, H.C.: An Extended Set of Haar-like Features for Bird Detection Based on AdaBoost. In: Kim, T.H., Adeli, H., Ramos, C. (eds.) Euro-Par 2011. CIP, vol. 206, pp. 160–169. Springer, Incorporated (2011).
10. Yan, L.V., Gong, Q.: Application of weighting 3D-Otsu method in image segmentation. Application Research of Computers. 28, 1576–1579 (2011).