Face Recognition Using Depth Images Base Convolutional Neural Network

Juxiang Chen
Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing
SouthWest University
Chongqing,China

Zhihao Zhang
Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing
SouthWest University
Chongqing,China

Liansheng Yao
Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing
SouthWest University
Chongqing,China

Bo Li
Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing
SouthWest University
Chongqing,China

Tong Chen*
Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing
SouthWest University
Chongqing,China
c_tong@swu.edu.cn

***Abstract*—Face recognition based on depth images is widely studied due to its advantages of 3 dimensional information and environment illumination insensitivity. The traditional recognition methods in this field mainly focus on hand-crafted feature design, which cannot achieve satisfactory result. In addition, there is no fixed face feature extraction method. To achieve a better face recognition performance on depth images, this paper proposes a method based on Convolutional Neural Networks(CNN). The experiment performed on database IIITD Kinect suggests that the proposed CNN architecture has better recognition performance than some traditional manual feature extraction methods, such as HOG and LBP.**

Keywords—face recognition; depth image; convolution neural network; deep learning

I. INTRODUCTION

With the rapid development of biometric identification technology, people have higher expectation on the accuracy of the identification [1]. Especially, face recognition has been widely applied in security system, video conference, and human-computer interaction [2]. The key step of various face recognition methods is to extract distinctive features and eliminate the influence of some inevitable factors in unconstrained environments, where face images have complex and large intra-personal variations, such as facial expressions, poses, decorations and illumination [3][4][5]. For face recognition applications in practical environments, uncontrolled illumination variety is a common and important handicap, which needs to be resolved. Since the RGB image is susceptible to illumination change, some alternative options have been used for the face recognition, such as depth images based on Kinect [6], near infrared (NIR) images [7], and hyperspectral images(HSI) [8][9].

The depth map is a characterization of the geometry of a face with grayscale values representing the distances of points on the face to the sensor. It is generated by using active infrared illumination. Therefore, we can perform face recognition under the insufficient ambient illumination condition or completely dark environment by using depth images. More importantly, a new generation of Kinect equipment has been greatly improved in the data acquisition speed, the use of the method has become simpler [10].

The traditional methods for face recognition using depth images often take the following steps: depth images are first preprocessed, and the features are then extracted or selected, the classifier is then trained based on the training set, finally the recognition rate is obtained on the test set. Ma et al [11] performed face recognition using depth images based on the covariance matrix method, and the results show that based on the information from depth images, a good recognition rate can be achieved. Amel et al [12] proposed an algorithm based on LBP, they extracted features from depth images and used k-Nearest Neighbor (KNN) as classifier. With a benchmark algorithm based on Eigenface [13], Yaser Sale and Eran Edirisinghe [14] carried out experiments in normal illumination and poor illumination environment, respectively. When images from the poor illumination are used for testing, the accuracy rate is 10% and 82.33% for normal images and depth-map images. The experimental results confirmed the advantage of depth maps in face recognition tasks under poor illumination.

For face recognition, an emerging alternative approach is the deep learning method [15][16], which has shown many advantages and has been applied in many different areas, such as computer vision [17] and pattern recognition [18].In this paper, we proposed a Convolutional Neural Networks (CNN) structure for face recognition based on depth image. The proposed CNN was tested on the IIITD_kinect database [19][20] .The CNN is a deep learning algorithm, which can automatically learn effective features from the original image data [21]. Compared to the traditional feature extraction algorithms, such as Local Binary Pattern (LBP), Histogram of Oriented Gradient (HOG) showed an average accuracy of 60% and 78%, the experimental results showed that the proposed CNN model achieved an average accuracy of 87%

978-1-7281-1374-6/19/$31.00 ©2019 IEEE

II. CONVOLUTION NEURAL NETWORK

A. Convolution Neural Network

Deep learning is an emerging research direction in the field of machine learning. It creates more abstract high-level features by combining low-level feature. As a kind of deep learning algorithms, CNN has a unique superiority in the field of speech recognition and image processing.

The basic structure of CNN consists of two layers: one layer is the feature extraction layer, i.e. the input of each neuron is connected with the local receptive field of the previous layer; the other layer is the feature mapping layer. Compared with the traditional recognition algorithms with complex feature extraction processes, CNN implements features extraction and classification into one step [22].

B. The architecture of network model

Several well-known network models have been widely applied, such as LeNet [23], Alexnet [24], GoogLeNet [25], and VGG [26]. However, if the images database is not large enough, a small-sized network may achieve slightly higher recognition rates than a large-scale network [27]. Therefore, we designedt a simple CNN model in order to achieve a good face recognition rate using depth images. The proposed network architecture is based on AlexNet [24]. It does not have the LRN layer and the fifth convolution layer, Compared to AlexNet.

The architecture of the proposed CNN model is shown in Figure 1.The network consists of a data input layer, four convolution layers, three pool layers and two fully connected layer with a softmax classifier. In addition, there are six layers appending Rectified Linear Units. Refer to the traditional Sigmoid activation function, it has inevitable malpractices that easily be saturated during the propagation process. While the Relu has a piecewise linear property, its prequel, posterior pass and derivative are all linear and easier to learn and optimize.

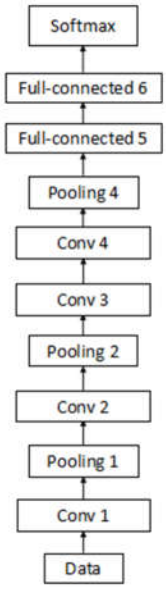

Fig. 1. The CNN architecture

The input of the proposed CNN model is the depth image. The depth image is convoluted with defined convolution kernels, where the convolution kernel size of the first convolution layer is 7×7 and the kernel size of the other convolution layers is smaller. The pooling layer is used for secondary feature extraction, the size of the all kernel function is 3×3, which can reduce the dimension and computation of the data. In addition, pooling layer can keep the invariance of image rotation, translation and expansion. Finally, softmax classification layer is used to classify the figure characteristics. We also use dropout, proposed by Hintion, to prevent over-fitting of the CNN model [28].

III. EXPERIMENT

A. Depth image database

There exist a few depth face recognition databases that are publicly available. Except IIITD Kinect face database, the maximum size of other databases is approximately 50 subjects.In order to implement face recognition experiments, a larger database is preferable. Therefore, we use the IIITD Kinect database.

IIITD Kinect face database comprises 106 male and female subjects with multiple depth images of each subject. The number of images per subject is variable with a minimum of 11 images and a maximum of 254 images. The total number of images in the database is 4605 and the size of each image is 640×480. Four images per subject are used as gallery. The remaining images from the database are used as probes. The detected depth faces include various faces in different expression, illumination, and pose.

Each fold contains gallery and probes, the experiment were carried out in five folds, and the gallery and probes in each fold are randomly selected from all the data collected by the database creator.

Because we use five times random subsampling for identification, the algorithm can be evaluated on each of the 5 folds individually. Meanwhile, we also perform face recognition tests on depth images using several common feature extraction algorithms, and finally give all the experimental results as well as the related analysis and discussion in Section C.

B. Parameter setting

The proposed network model has 11 layers. The size of the input depth image is 227×227. Convolution layer 1 (Conv 1 in Fig 1) employs 96 different convolution kernels whose output image is 111×111 in size. The output data of max pooling layer 1 (Pooling 1 in Fig 1) is decreased into 55×55, which undertakers down-sampling. Then, the data are convoluted and pooled by Conv 2 and Pooling 2, whose dimension and size become 256 and 27×27. After two convolution layers (Conv 3 and Conv 4) and the one pooling layer (Pooling 4), the feature extraction of the depth image is completed. Finally, the dropout technique is used to prevent the over-fitting of the model in the two fully connected layers (Full-connected 5 (fc5) and Full-connected 6 (fc6) in Fig 1). The all-connected layer fc7 in the softmax classifier is used to achieve classification and identification. The data are grouped into 106 classes. The weight decay is utilized to impose a penalty on the loss function, making the learning algorithm more inclined to get the model with small weight. The structure and the output size of each layer are summarized in Table I, where *s* denotes step length and *p* denotes padding.

TABLE I. PROPOSED CNN STRUCTURE AND THE OUTPUT SIZE

layer	size	output
data		227×227
Conv 1	7×7 s2 p0	111×111×96
Pool 1	3×3 s2 p0	55×55×96
Conv 2	5×5 s1 p2	55×55×256
Pool 2	3×3 s2 p0	27×27×256
Conv 3	3×3 s1 p1	27×27×512
Conv 4	3×3 s1 p1	27×27×512
Pool 4	3×3 s2 p0	13×13×512
Fc 5		4096
Fc 6		4096
Softmax		106

With the purpose of proving the validity of the CNN, we compare the proposed CNN method with other state-of-art algorithms, such as LBP, HOG.The parameters of the two algorithms were set based on the references [29][30].

C. Experimental results and discussion

We used five-fold cross validation to test LBP, HOG, and CNN. The LBP and HOG were combined with k-Nearest Neighbor (KNN) classifier for classification. Since our network architecture was based on AlexNet, we also tested the performance of AlexNet. The experimental results are given in the table Ⅱ. Considering that noise could exist in real application, we add Gaussian noise with mean of 0 and variance of 0.01 to the depth images, and also applied all methods on the noise-added images. This experimental results are given in the Table Ⅲ.

TABLE II. FACE RECOGNITION RATE % (RANK-5) IN FIVE EXPERIMENTS

Algorithm	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
LBP	60.82	61.42	60.27	58.74	59.63
HOG	79.29	78.16	78.14	76.72	79.86
Alexnet	85.81	83.62	81.51	83.15	83.31
Proposed	88.76	87.45	85.04	88.15	88.45

TABLE III. FACE RECOGNITION RATE % (RANK-5) IN FIVE EXPERIMENTS (GAUSSIAN NOISE)

Algorithm	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
LBP	25.78	24.25	25.66	26.76	23.75
HOG	36.74	34.21	39.58	39.46	38.08
Alexnet	85.62	84.25	82.12	84.37	82.87
proposed	88.12	86.56	83.62	86.43	87.12

From Table Ⅱ, we can see that HOG has better recognition performance than LBP. This result accord with other publications.

AlexNet can achieve higher recognition rate than traditional methods (LBP and HOG), and the proposed CNN can achieve highest recognition rate. This results suggest that CNN is a good way for depth face recognition.

The proposed CNN is a simplified version of AlexNet, and has better recognition performance than AlexNet. This may prove that the special designed network is more suitable for the IIITD Kinect database, which is a relatively small database.

It is observed from Table III that the CNN are much more robust than traditional methods when noise is present in the depth images. The recognition rate of LBP and HOG drops sharply (decrease of 35% to 45%), However, the recognition rate achieved by CNN based methods were hardly decreased. The proposed CNN under noise condition can achieve highest recognition rate as well.

The CNN proposed in this paper is relatively shallow for adapting to the small database. With the increase of the number of layers, the recognition rate might be increased, but could reach a top and then decrease. In addition, the recognition rate based on CNN method could be slightly improved by using data augmentation, which can increase the database by generating synthetic images.

IV. CONCLUSION

Depth image is independent of ambient illumination change and is more practical in real application. In this paper, we presented a CNN architecture for face recognition using depth images. The CNN is special designed for small database, and can achieve recognition rate of 86%. It is also robust to the noise environment, and hardly affected by the Gaussian noise. Compared to traditional depth face recognition method, the proposed CNN is more suitable in practical application, both in terms of better recognition rate and robustness.

ACKNOWLEDGMENT

We would like to thank the support from the National Natural Science Foundation of China (Grant No. 61301297) and Southwest University Doctoral Foundation (No. SWU115093).

REFERENCES

[1] N. Kshetri, "Information and communication technologies, strategic asymmetry and national security," Journal of International Management, 2005, pp: 563-580.

[2] R. Chellappa, C. L. Wilson, S Sirohey, "Human and machine recognition of faces: A survey," Proceedings of the IEEE, 1995, 83(5), pp: 705-741.

[3] Y Adini, Y Moses, S Ullman, "Face recognition: The problem of compensating for changes in illumination direction," IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7), pp: 721-732.

[4] A Wagner, J Wright, A Ganesh, Z Zhou, H Mobahi, Y Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation, " IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(2), pp: 372-386.

[5] T Zhang, Y. Y. Tang, B Fang, Z. W. Shang, X. Y. Liu, "Face recognition under varying illumination using gradientfaces," IEEE Transactions on Image Processing, 2009, 18(11), pp: 2599-2606.

[6] B. Y. L. Li, A. S. Mian, W. Liu, A. Krishna, "Using kinect for face recognition under varying poses, illumination and disguise," Application of Computer Vision(WACV), 2013 IEEE Workshop on. IEEE, 2013, pp: 186-192.

[7] B. Klare, A. K. Jain, "Heterogeneous face recognition: Matching nir to visible light images" Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE, 2010, PP: 1513-1516.

[8] Z. Pan, G. Healey, M. Prasad, B. Tromberg, "Face recognition in hyperspectral images," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(12), pp: 1552-1560.

[9] Z. Pan, G. E. Healey, B. Tromberg, "Multiband and spectral eigenfaces for face recognition in hyperspectral images," Defense and Security. International Society for Optical and Photonics, 2005, pp: 144-151.

[10] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced Computer Vision With Microsoft Kinect Sensor: A review," IEEE Transactions on Cybernetics, 2013, 43(5), pp: 1318-1334.

[11] J. H. Ma, H. Zhang, Q. Ji, "Robust face recognition using block based on covariance matrix to represent depth image sets," Application Research of Computers, 2016.33(12)

[12] A. Aissaoui, J. Martinet, C. Djeraba, "Dlbp: A novel descriptor for depth image based face recognition," Image Processing (ICIP), 2014 IEEE International Conference on. IEEE, 2014, pp: 298-302.

[13] M. Turk, A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neurosciences, vol.3, no.1, pp.71-86, 1991.

[14] Y. Saleh, E. Edirisinghe, "Novel approach to enhance face recognition using depth maps," Signals and Image Processing (IWSSIP), 2016 International Conference on. IEEE,2016, pp: 1-4

[15] Z. Wu, M. Peng, T. Chen, "Thermal face recognition using convolutional neural network," Optoelectronics and Image Processing (ICOIP), 2016 International Conference on, IEEE, 2016, pp: 6-9.

[16] X. Zhang, M. Peng, T. Chen, "Face recognition from near-infrared images with convolutional neural network," Wireless Communication and Signal Processing (WCSP), 2016 8th International Conference on. IEEE, 2016, pp: 1-5

[17] Y. Bengio, "Learning deep architectures for AI," Foundations and trends® in Machine Learning, 2009, 2(1), pp: 1-127

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, 2015,521(7553), pp: 436-444.

[19] G. Goswami, S. Bharadwaj, M. Vatsa, and R. Singh, "On RGB-D face recognition using Kinect," Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on. IEEE, 2013, pp.1-6.

[20] G. Goswami, M. Vatsa, and R. Singh, "RGB-D face recognition with texture and attribute features," IEEE Transactions on Information Forensics and Security, 2014, 9(10), pp: 1629-1640.

[21] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks, 2015, 61, pp: 85-117.

[22] G. Chen, D. Clarke, M. Giulianli, A. Gaschler, A. Knoll, "Combining unsupervised learning and discrimination for 3D action recognition," Signal Processing, 2015, 110, pp: 67-81.

[23] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, 1998,86(11)86, pp: 2278‑2324.

[24] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, 2012, pp: 1097-1105.

[25] C. Szegedy, W. Liu, Y. Q. Jia, P. Sermanet, S. Reed, D. Anguelow, D. Erhan, "Going deeper with convolutions," Conference on Computer Vision and Pattern Recognition, 2015, pp: 1-9.

[26] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv: 1409. 1566. 2014.

[27] M. Peng, C. Y. Wang, T. Chen, G. Y. Liu, "NIRFaceNet: A convolutional neural network for near-infrared face identification," Information 2016, 7(4): 61.

[28] G. E. Hinton, N Srivastava, A Krizhevsky, I Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," arXiv preprint arXiv: 1207.0580, 2012.

[29] Ruiz-del-Solar, Javier, "Thermal Face Recognition in Unconstrained Environments Using Histograms of LBP Features," Local Binary Patterns: New Variants and Applications. Springer Berlin Heidelberg, 2014, pp. 219-243

[30] Hermosilla, Gabriel, "Study of Local Matching-Based Facial Recognition Methods Using Thermal Infrared Imagery," International Journal of Pattern Recognition and Artificial Intelligence 29.08, 2015, 1556012.