

A Multi-Scale Parallel Convolutional Neural Network Based Intelligent Human Identification Using Face Information

Chen Li*, Mengti Liang*, Wei Song*, and Ke Xiao*

Abstract

Intelligent human identification using face information has been the research hotspot ranging from Internet of Things (IoT) application, intelligent self-service bank, intelligent surveillance to public safety and intelligent access control. Since 2D face images are usually captured from a long distance in an unconstrained environment, to fully exploit this advantage and make human recognition appropriate for wider intelligent applications with higher security and convenience, the key difficulties here include gray scale change caused by illumination variance, occlusion caused by glasses, hair or scarf, self-occlusion and deformation caused by pose or expression variation. To conquer these, many solutions have been proposed. However, most of them only improve recognition performance under one influence factor, which still cannot meet the real face recognition scenario. In this paper we propose a multi-scale parallel convolutional neural network architecture to extract deep robust facial features with high discriminative ability. Abundant experiments are conducted on CMU-PIE, extended FERET and AR database. And the experiment results show that the proposed algorithm exhibits excellent discriminative ability compared with other existing algorithms.

Keywords

Face Recognition, Intelligent Human Identification, MP-CNN, Robust Feature

1. Introduction

Biometric based human identification has been widely studied in the field of artificial intelligence and pattern recognition for years. Now ranging from Internet of Things (IoT) application, interactive multi-media [1], self-service bank, intelligent surveillance to public safety, access control and information security, biometric based human identification has been introduced as a key procedure to improve the intelligent degree or strengthen the security of the applications mentioned above.

The most widely used biometric feature for human identification method have been fingerprint recognition nowadays. It has been applied to attendance system, unlock the smart phone and laptop, access control and payment. However, its disadvantages are becoming obvious too. Since fingerprint recognition requires contact with the sensor, it cannot be used to long-distance or none-intrusive applications. Besides, it can be easily invalidated with grimy or wet hand and can be easily cheated by a

* This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received September 7, 2018; accepted October 22, 2018.

Corresponding Author: Ke Xiao (zehan_xiao@163.com)

* School of Computer Science, North China University of Technology, Beijing, China (lichen@ncut.edu.cn, 759031154@qq.com, songwei@ncut.edu.cn, zehan_xiao@163.com)

fake fingerprint film.

Compared with this, human identification based on face images appears to be more convenient, stable and reliable, because face images can be captured from a distance without contact or cooperate of human, and it is not easy to fake. However early face recognition researches usually apply high-resolution frontal face images [2,3] database captured under constrained circumstance, which can help achieve high recognition rate but cannot meet the requirement of wider applications. Since 2D face images are usually captured from a long distance in an unconstrained environment [4], face recognition under complex condition has become the new research hotspot. It can help exploit the advantage and make human recognition appropriate for wider applications with higher security and convenience.

The key difficulties here include gray scale change caused by illumination variance, occlusion caused by glasses, hair or scarf, self-occlusion and deformation caused by pose variation. To conquer these, many solutions have been proposed. However, most of them only improve recognition performance under one influence factor, which still cannot meet the real face recognition scenario. In this paper we propose a multi-scale parallel convolutional neural network (MP-CNN) architecture to solve the face recognition problem in complex environment. This paper is organized as follows: Section 2 reviews and discusses the related work. Section 3 describes the novel paralleled CNN architecture proposed in this paper. In Section 4 presents the experimental results on different conditions and the comparison with other state-of-the-art methods to verify the face recognition efficiency of the CNN structure proposed by this paper in the complex environment. And the conclusions are drawn in Section 5.

2. Related Works

To combat the illumination variation which exist widely in real scenario face recognition, the most commonly used approaches are preprocessing and normalization techniques such as: traditional gray scale processing methods, the histogram equalization [5], wavelet based image fusion [6], etc. However, these kinds of methods can only solve slightly illumination variation. Other widely used approaches are reflectance model based method, however the modeling and optimizing process are very complex. Researchers try to extract features that are robust to gray scale and appearance change cause by illumination variation. Hence, image filters which are applied on the whole face (holistic) or local face area are discussed. The holistic methods including Principal Component Analysis (PCA) [7], Linear Discriminant Analysis (LDA), and Information Discriminant Analysis (IDA) [7] have been fully explored, which are proven to be non-robust to grays scale change caused by illumination variation. Local features including local binary patterns (LBP) [8], center-symmetric local binary patterns (CSLBP) [9], local directional number pattern (LDN) [10,11], and dense sampling based local binary patterns (DSLBP) [12], show much better [13] discriminative ability and are able to accommodate local variation. However, the performances of local features are usually sensitive to smooth regions, which is an obvious drawback while expressing face images.

For the face recognition with occlusion, sub-space based methods have been widely studied, including dual-kernel based face recognition method [14], two-dimensional fisher discriminant analysis (K2DFDA) [15], etc. Besides, sparse representation-based classification (SRC) has led to the state-of-the-art performance in occluded face recognition, such as the non-negative sparse representation based

general classification algorithm [16] and the occlusion dictionary based method [17]. However, these methods normally need to make certain assumptions and the computational complexity are high. Also, algorithms mentioned above usually concentrated on only one influence factor instead of multiple influence factors which exists in real scenario.

To compensate these disadvantages, researchers try to exact more robust facial features with higher distinguished ability. The rapid progress achieved in deep network especially in CNN provide a novel and feasible approach. CNN is a feedforward deep neural network, which is inspired by the structure of biological neural networks and visual systems. It has shown obvious advantages in image classification and recognition [18-21]. However traditional CNNs cannot achieve satisfied performance in face recognition under complex environment, including illumination variation, pose variation or partial occlusion. Thus, more complex network structures are implemented, including DeepFace [22] proposed by Facebook, VGG [23], DeepId [24], and Google's FaceNet [25,26]. They all achieved state-of-the-art performance. The common characters of these representative CNNs are very deep and sophisticated structure, complex parameter tuning and the dependence on huge scale data sets, which make the computational complexity and the requirement of hardware are both very high.

Hence, in this paper, we propose a MP-CNN to extract deep robust facial features with high discriminative ability as well as much lower computational complex compared with the CNNs mentioned above. And through a multi-scale and parallel structure, deep features in different scale can be extracted and fused for face recognition, which compensate the shortage of traditional single CNN or simple parallel CNNs.

3. Technical Approach

3.1 Architecture

In this section, the proposed MP-CNN architecture will be detailed. CNN is a deep neural network which usually consist of multiple convolution layer to extract the deep features. Besides adding the pooling layer is a feasible way to reduce the dimension of the feature map. Since traditional CNNs cannot meet the requirement of face recognition under complex environment, including illumination variation, pose variation or partial occlusion. Researchers begin to study the more complex and deeper CNN network structure. To express the face image in different scale and abstract the deep and robust features, a parallel convolutional neural network with four different CNNs is proposed. The MP-CNN structure proposed in this paper is shown in the Fig. 1.

As shown above, the proposed MP-CNN is composed by four CNN networks in a cascaded form. The four CNN networks are named as CNN-11×11, CNN-7×7, CNN-5×5, and CNN-3×3 from top to bottom. Each of them separately has three convolutional layers followed with a pooling layer. To fully express the face image, the convolution kernels of these four CNN networks have different kernel scale. The size of the convolution kernels are 11×11, 7×7, 5×5, 3×3 from top to bottom. The output of each CNN's third pooling layer combines into a four-channel image before been feed into the fully connection layer.

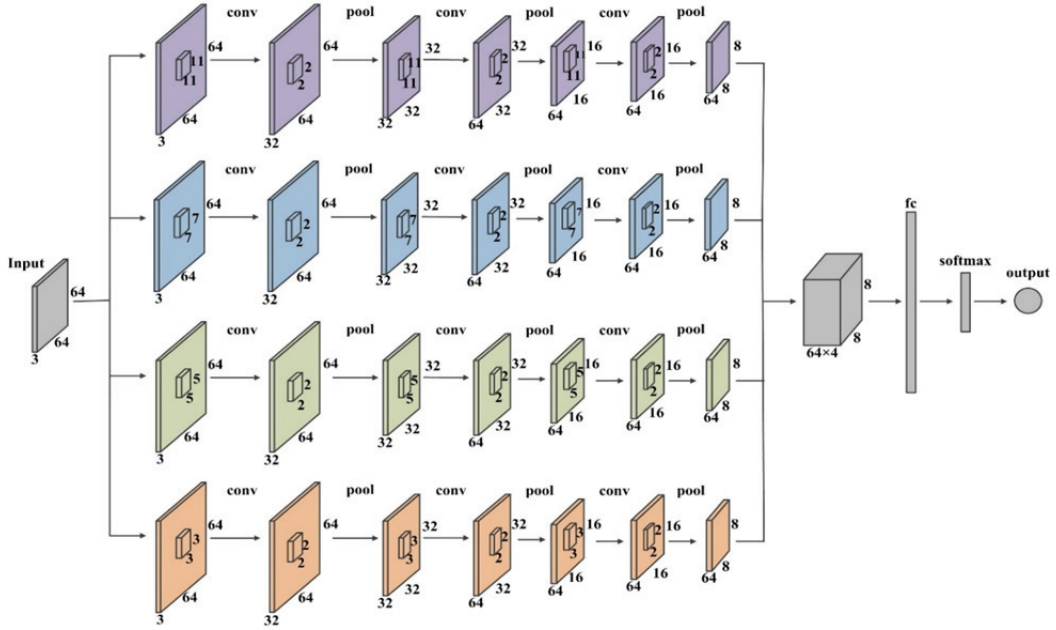


Fig. 1. The structure of MP-CNN.

3.2 Implementation Details

3.2.1 Multi-core convolution and pooling layers

For the convolutional layer, the network proposed in this paper uses multi-core convolution. Since that only one feature map is not sufficient to fully reflect all the distinguished information of the face image, it is necessary to choose different convolution kernels to acquire multi-scale features of the image, so as to obtain multiple feature maps of the original images. Four different convolution kernel sizes are set for four parallel CNN networks. The size of the convolution kernels are 11×11 , 7×7 , 5×5 , 3×3 from top to bottom. For each individual CNN network, its three convolutional layers remains the same size.

The neuron number in the hidden layer is related to the size of the original image, the size of the convolution kernel, and the step size of the convolution kernel in the image. After the feature maps are obtained by the convolution operation, they can be used as an input to train the classifier.

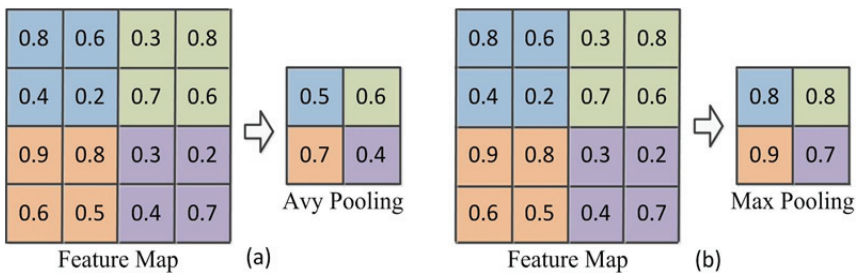


Fig. 2. Average pooling (a) and max pooling (b).

However, the dimension of the feature vector obtained after the convolution operation is still very high, which can easily cause overfitting of the classifier. In order to solve this problem, pooling layer is applied behind each convolution layer. Pooling can be seen as a feature selection procedure, which effectively reduces the feature dimension as well as the network parameters amount. There are usually two pooling strategy including average pooling and max pooling, as shown in Fig. 2.

In this proposed structure, the maximum pooling is applied to each pooling layer, and the size is 2×2 with a stride of one.

3.2.2 Softmax layer

As shown in Fig. 1, after the last pooling layer, all the feature maps are jointed together and fed into the full connection layer, which is followed by the softmax layer. The output of softmax layer is a probability distribution, which makes it more suitable for probabilistic interpretation in classification tasks compared with selecting one maximum value.

The softmax layer in this paper applies the cross entropy loss function, which is defined as follow:

$$L = - \sum_{k=1}^n \sum_{i=1}^C t_{ki} \log(p_{ki}). \quad (1)$$

As shown in Eq. (1), $\log(\cdot)$ is a logarithmic function, t_{ki} can be seen as the target probability distribution, and p_{ki} is the output of the estimated label probability distribution.

For a single specific sample, its cross entropy loss can be expressed as follow:

$$l_{CE} = - \sum_{i=1}^C t_i \log(p_i) \quad (2)$$

As shown, t_i is the real category label. And p_i is the predicted probability of the specific sample belonging to category i , which can be expressed with softmax function, as shown in Eq. (3).

$$p_i = \frac{e^{m_i}}{\sum_{k=1}^C e^{m_k}} \quad \forall i \in 1 \dots C \quad (3)$$

The objective of training deep network using the constructed loss function is to make the estimated label probability distribution as close as the target probability distribution. Hence, the derivative or gradient of the loss function need to be calculated and passed back to the previous layer during back-propagation. For a single sample, the derivative of the loss function on the input m_j can be calculated as follow:

$$\frac{\partial l_{CE}}{\partial m_j} = - \sum_{i=1}^C \frac{\partial t_i \log(p_i)}{\partial m_j} = - \sum_{i=1}^C t_i \frac{\partial \log(p_i)}{\partial m_j} = - \sum_{i=1}^C t_i \frac{1}{p_i} \frac{\partial p_i}{\partial m_j} \quad (4)$$

Apparently, $\frac{\partial p_i}{\partial m_j}$ is the derivative of the softmax function on the input m_j . The derivative result are usually discussed under two situation, including:

$$\begin{aligned} \text{when } i=j: \frac{\partial p_i}{\partial m_j} &= p_i(1 - p_j) \\ \text{when } i \neq j: \frac{\partial p_i}{\partial m_j} &= -p_i p_j \end{aligned} \quad (5)$$

Hence, the derivative of the loss function is:

$$\frac{\partial l_{CE}}{\partial m_j} = p_j - t_j \quad (6)$$

The equation shown above is quite concise to illustrate the idea that when conducting optimization through minimizing the derivative of loss function, the estimated label will be as close as the real category label. Compare with the minimum mean-square error loss function, the cross-entropy loss function used in this paper has a less flat area, so it makes the training process easier to escape from the local minimum point and achieves much better training efficiency.

4. Experiments and Analysis

In this section, extensive experiments are conducted to verify the performance of the proposed method in complex condition, including face recognition under illumination variation, expression variation, pose variation and partial occlusion. Three different databases with multiple interference factors are adopted in this paper. For each dataset, 20% of all images are randomly selected as the test set, while the remaining images are used as training set. A total of 150 epoch of iterative training are performed with a learning rate of 0.0001. To fully verify the effectiveness of our proposed MP-CNN, many comparison experiments are conducted using other algorithms, including the single CNN (1-CNN) as well as simple parallel CNN (4-CNN) which is constructed with four same single CNNs without the multi-scale concept. The traditional single CNN structure is shown in Fig. 3, and 4-CNN structure is shown in Fig. 4. Besides, the proposed method is also compared with the renowned LBP algorithm and the CSLBP algorithm.

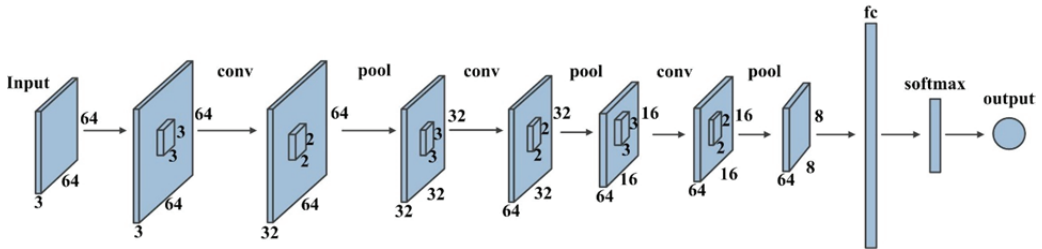


Fig. 3. 1-CNN structure.

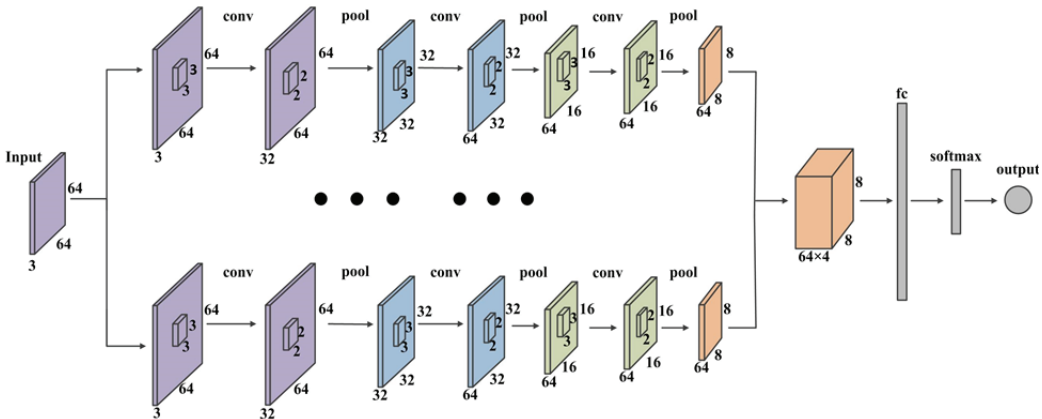


Fig. 4. 4-CNN structure.

4.1 Experiments on CMU-PIE Database

To verify the face recognition performance of the proposed MP-CNN under multi-pose as well as severe illumination variation conditions, the CMU-PIE face database are used for experiment. The CMU-PIE face database includes 40,000 photos from 68 people, in which it contains 13 poses of each person, as well as 43 illumination conditions and 4 expressions. Hence CMU-PIE is the most commonly used database for research on multi-pose face recognition under illumination variation. In this paper we randomly select 68 people with 170 face images per person for experiments. The total number is 11,560 images, during which 2,312 images are used for test and other 9,248 images are applied for training. There is no overlap between the test and training set. Example images of the CMU-PIE face database are shown in Fig. 5.

Face recognition are conducted on the CMU-PIE face database using our proposed MP-CNN as well as other four methods including 1-CNN, 4-CNN, LBP, and CS-LBP. The CMC curves of the five methods are shown in Fig. 6.

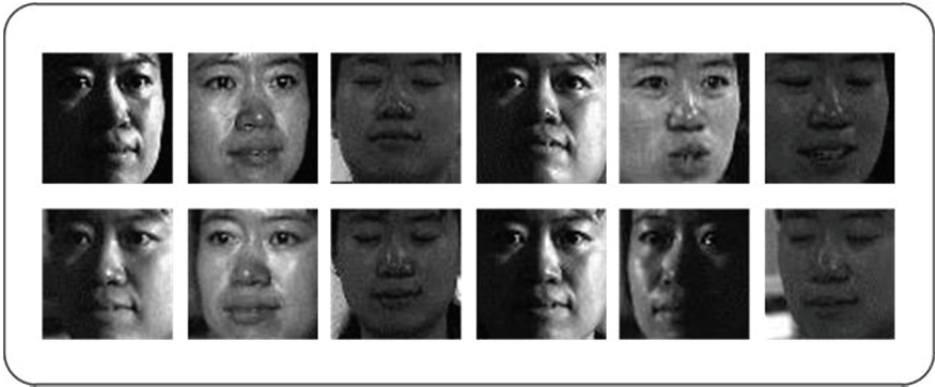


Fig. 5. Image examples of CMU-PIE face database.

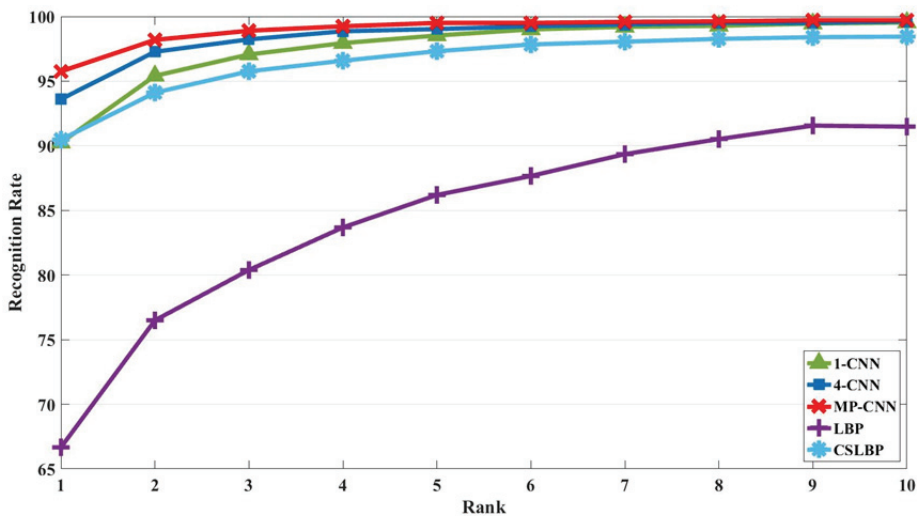


Fig. 6. Experiment results comparison on CMU-PIE database.

In order to show the contrast more clearly, the RANK1 recognition rate of each algorithm is listed in Table 1.

Table 1. The RANK1 recognition rates on CMU-PIE face database

Method	Accuracy (%)
MP-CNN	96.61±0.9
4-CNN	93.37±0.3
1-CNN	91.86±0.6
LBP	66.67
CSLBP	90.48

As shown, the recognition rate of the proposed method in this paper reaches 96.61%, while the recognition rate of the 4-CNN and 1-CNN are 93.37% and 91.86% separately. The recognition rate of LBP is 66.67%, and CSLBP is 90.48%. It can be seen that compared with 1-CNN structure, 4-CNN, LBP and CS-LBP, the MP-CNN structure proposed in this paper achieves the best recognition rate on face recognition under sever illumination variation and slight pose and expression variation.

4.2 Experiments on Extended FERET Database

In the second experiment, the FERET face database are applied for face recognition under illumination variation. The original FERET face database include more than 10,000 photos from more than 1,000 people with different expressions, light conditions, postures and ages. 200 subjects are randomly selected for this experiment. To further discuss the robustness of our proposed method on different illuminations, the database is augmented by transforming the overall illumination of the images. The extended data set are consisted of 140 images per person, and the total account is 28,000 images. 22,400 images are randomly selected for training, and 5,600 images are remained for test. There's no overlap between the training set and the test set. Examples of extended FERET database are shown in Fig. 7.

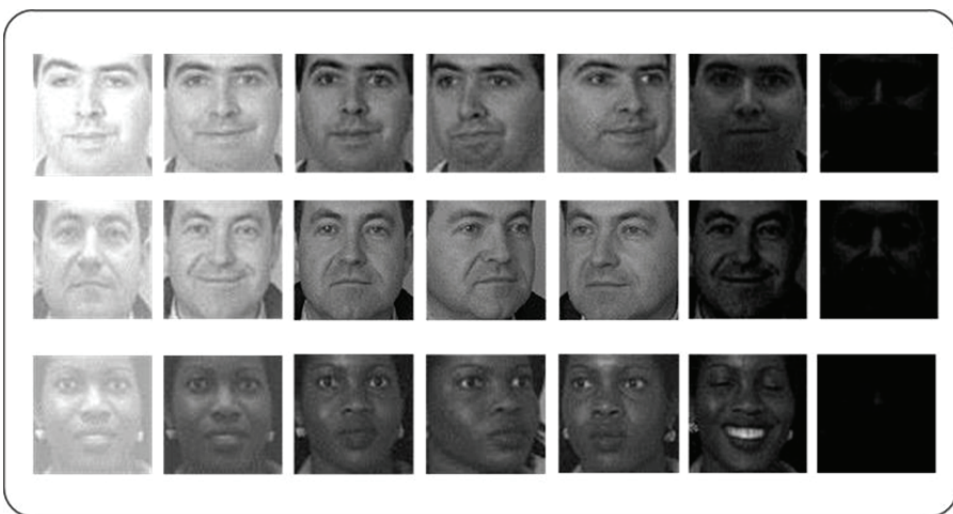


Fig. 7. Image examples of the extended FERET face database.

The recognition performance of the five methods on extended FERET database are shown in Fig. 8.

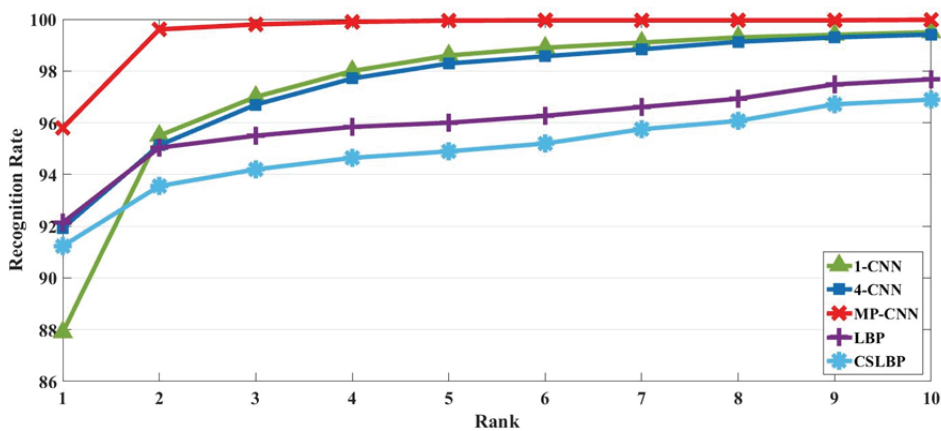


Fig. 8. Experiment results comparison on the extended FERET database.

Table 2. The RANK1 recognition rates on extended FERET face database

Method	Accuracy (%)
MP-CNN	95.48±0.6
4-CNN	91.97±0.7
1-CNN	87.93±0.3
LBP	92.13
CSLBP	91.23

In order to show the contrast more clearly, the RANK1 recognition rate of each algorithm is listed in Table 2.

As shown, the recognition rate of the proposed method is 95.48%, and the recognition rate of the 4-CNN and 1-CNN are 91.97% and 87.93%. Besides, the recognition rates of LBP and CSLBP are 92.13%, and 91.23% separately. It can be seen that compared with the 1-CNN and 4-CNN, the local feature based algorithms achieve better performance which is different from the first experiment. The main reason is that the database applied in this experiment is augmented by illumination transfer, which don't bring in much new information virtually. That is to say, the main features are still from the original images of the database. So, this can be seen as another kind of small sample problem. The experimental results shown that for this kind of small sample problem, traditional single CNN as well as the simple parallel CNN cannot achieve satisfying performance, even not as good as local feature based algorithms. Compared with the four algorithms, the MP-CNN structure proposed in this paper shows the best performance in the augmented FERET database, which effectively verified the effectiveness of the MP-CNN.

4.3 Experiments on Extended AR Database

In the third experiment, to further verify the performance of our proposed MP-CNN algorithm for face recognition under severe occlusion as well as facial expression, the AR face database are adopted.

AR database contains 3,288 images of 116 people, including illumination and expression changes, as well as partial occlusion caused by wearing glasses and beards. We randomly select 100 people for this experiment. The database is extended by randomly adding occlusion blocks. The size of the occlusion block is 10% of the face image. Through this the database is extended to 140 images per person, and a total of 14,000 images. Among them, there are 2,800 test images and 11,200 train images, and no overlap between the training set and the test set. The examples of the extended database are shown in Fig. 9.

The recognition performance of the five methods on expanded AR database are shown in Fig. 10.



Fig. 9. Image examples of the extended AR face database.

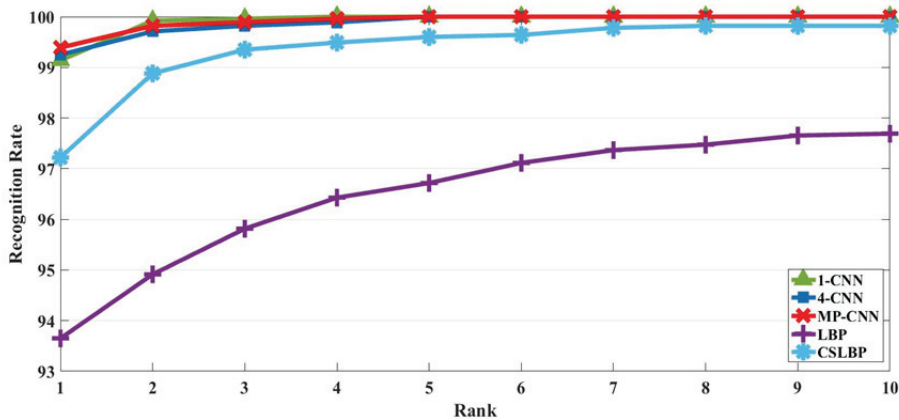


Fig. 10. Experiment results comparison on the extended AR database.

In order to show the contrast more clearly, the RANK1 recognition rate of each algorithm is listed in Table 3.

Recognition performance of our proposed algorithm with other four algorithms including single CNN, 4-CNN, LBP as well as CSLBP are compared. The experimental schemes are the same, and the comparison of experimental results are demonstrated in Fig. 10. As shown, for partial occluded face

recognition, our proposed method achieves a 99.46% RANK1 recognition rate. The 1-CNN and 4-CNN achieve 98.15% and 98.87% recognition rate, respectively. The recognition rate of LBP is 93.65% while the recognition rate of CSLBP is 97.12%. In order to show it more clearly, the RANK1 preferred recognition rate of each algorithm is listed in Table 3. It can be seen that the proposed MP-CNN shows the best recognition performance for face recognition with partial occlusion as well as facial expression.

Table 3. The RANK1 recognition rates on the enhanced AR database

Method	Accuracy (%)
MP-CNN	99.46±0.2
4-CNN	98.87±0.5
1-CNN	98.15±0.4
LBP	93.65
CSLBP	97.12

5. Conclusion

With the rapid development of Artificial Intelligence technology, IoT application, intelligent self-service bank, intelligent surveillance of public safety and access control have been an indispensable part of human daily life. Almost all these applications require intelligent human identification to improve its security and user experience. Among all the biometric features which can be applied to human identification, face image has natural advantages for the above mentioned intelligence application since it can be captured from a long-distance without contact or cooperate of human. However, gaps lying between face recognition in lab condition and in real word are still inevitable. The key difficulties are gray scale change caused by illumination variance, occlusion caused by glasses, hair or scarf, self-occlusion and deformation caused by pose variation. To conquer these, we propose a multi-scale parallel CNN architecture consist of four multi-scale CNN structure, through which deep features in different scale can be extracted and fused for face recognition. The shortage of traditional single CNN or simple parallel CNNs can be compensated through the multi-scale and parallel structure. Abundant experiments are conducted under different complex recognition condition including illumination and slight pose variation, illumination variation and partial occlusion. The comparison results of the proposed method with four existing renowned algorithms show the effectiveness of the proposed MP-CNN.

Acknowledgement

This paper is supported by the National Key R&D Program of China (No. 2017YFB0802300), Research Project of Beijing Municipal Education Commission (No. KM201810009005), the North China University of Technology “YuYou” Talents support Program, the Beijing Young Topnotch Talents Cultivation Program, the Beijing Talents Support Program (Backbone Talent Program), High Innovation Program of Beijing (No. 2015000026833ZK04), NCUT “Science and Technology Innovation Engineering Project.”

References

- [1] W. Song, G. Sun, S. Fong, and K. E. Cho, "A real-time infrared LED detection method for input signal positioning of interactive media," *Journal of Convergence*, vol. 7, article ID. 16071002, 2016.
- [2] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 large-scale experimental results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 831-846, 2010.
- [3] P. J. Grother, G. W. Quinn, and P. J. Phillips, "Report on the evaluation of 2D still-image face recognition algorithms," National Institute of Standards and Technology, *NIST Interagency Report No. 7709*, 2010.
- [4] A. Moeini, K. Faez, and H. Moeini, "Unconstrained pose-invariant face recognition by a triplet collaborative dictionary matrix," *Pattern Recognition Letters*, vol. 68, pp. 83-89, 2015.
- [5] K. Ramirez-Gutierrez, D. Cruz-Perez, J. Olivares-Mercado, M. Nakano-Miyatake, and H. Perez-Meana, "A face recognition algorithm using eigenphases and histogram equalization," *International Journal of Computers*, vol. 5, no. 1, pp. 34-41, 2011.
- [6] S. U. Khan, W. Y. Chai, C. S. See, and A. Khan, "X-ray image enhancement using a boundary division wiener filter and wavelet-based image fusion approach," *Journal of Information Processing Systems*, vol. 12, no. 1, pp. 35-45, 2016.
- [7] Z. Nenadic, "Information discriminant analysis: Feature extraction with an information-theoretic objective," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1394-1407, 2007.
- [8] L. Lei, D. H. Kim, W. J. Park, and S. J. Ko, "Face recognition using LBP eigenfaces," *IEICE Transactions on Information and Systems*, vol. 97, no. 7, pp. 1930-1932, 2014.
- [9] J. Li, Y. Zhao, and D. Quan, "The combination of CSLBP and LBP feature for pedestrian detection," in *Proceedings of 2013 3rd International Conference on Computer Science and Network Technology (ICCSNT)*, Dalian, China, 2013, pp. 543-546.
- [10] A. R. Rivera, J. R. Castillo, and O. O. Chae, "Local directional number pattern for face analysis: face and expression recognition," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1740-1752, 2013.
- [11] A. R. Rivera and O. Chae, "Spatiotemporal directional number transitional graph for dynamic texture recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 2146-2152, 2015.
- [12] J. Ylioinas, A., Hadid, Y. Guo, and M. Pietikainen, "Efficient image appearance description using dense sampling based local binary patterns," in *Computer Vision-ACCV 2012*. Heidelberg: Springer, 2012, pp. 375-388.
- [13] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: a comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803-816, 2009.
- [14] X. Z. Liu and H. W. Ye, "Dual-kernel based 2D linear discriminant analysis for face recognition," *Journal of Ambient Intelligence and Humanized Computing*, vol. 6, no. 5, pp. 557-562, 2015.
- [15] X. Z. Liu, P. S. Wang, and G. C. Feng, "Kernel-based 2D fisher discriminant analysis with parameter optimization for face recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 27, no. 8, article no. 1356010, 2013.
- [16] B. Zhang, Z. Mu, C. Li, and H. Zeng, "Robust classification for occluded ear via Gabor scale feature-based non-negative sparse representation," *Optical Engineering*, vol. 53, no. 6, article no. 061702, 2013.
- [17] L. Yuan, W. Liu, and Y. Li, "Non-negative dictionary based sparse representation classification for ear recognition with occlusion," *Neurocomputing*, vol. 171, pp. 540-550, 2016.
- [18] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 3642-3649.

- [19] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. LeCun, "Pedestrian detection with unsupervised multi-stage feature learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, 2013, pp. 3626-3633.
- [20] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," in *Proceedings of 2012 21st International Conference on Pattern Recognition (ICPR)*, Tsukuba, Japan, 2012, pp. 3304-3308.
- [21] P. Luo, X. Wang, and X. Tang, "Hierarchical face parsing via deep learning," in *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 2480-2487.
- [22] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "DeepFace: closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 2014, pp. 1701-1708.
- [23] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proceedings of the British Machine Vision Conference (BMVC)*, Swansea, UK, 2015.
- [24] Y. Sun, D. Liang, X. Wang, and X. Tang, "DeepID3: face recognition with very deep neural networks," 2015 [Online]. Available: <https://arxiv.org/abs/1502.00873>.
- [25] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: a unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, 2015, pp. 815-823.
- [26] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, 2015, pp. 1-9.



Chen Li <https://orcid.org/0000-0001-5983-5895>

She received B.S. degrees from University of Science and Technology Beijing in 2001, and received her Ph.D. degree in 2013 from University of Science and Technology Beijing. She is currently an associate professor in North China University of Technology, Beijing, China. Her research interests include image processing, pattern recognition and 3D reconstruction.



Mengti Liang <https://orcid.org/0000-0002-4299-5128>

She received B.S. degree from North China University of Technology in 2017. She is currently a postgraduate student in North China University of Technology. Her current research interests include image processing and pattern recognition.



Wei Song <https://orcid.org/0000-0003-0649-8850>

He received her Ph.D. degree in 2013 from University of Science and Technology Beijing. He is currently a professor in North China University of Technology, Beijing, China. His research interests include data mining and analysis, big data analysis.



Ke Xiao <https://orcid.org/0000-0002-8654-1339>

He received B.S. degree from Jilin University in 2002, and received M.S. degree from Nankai University in 2005. He received the Ph.D. degree from Beijing University of Posts and Telecommunications in 2008. He is currently an associate professor at School of Computer Science and North China University of Technology. His main research interests include communication security and pattern recognition.