

A Novel Face Recognition Algorithm based on the Deep Convolution Neural Network and Key Points Detection Jointed Local Binary Pattern Methodology

Wen-zhun Huang[†] and Shan-wen Zhang*

Abstract – This paper presents a novel face recognition algorithm based on the deep convolution neural network and key point detection jointed local binary pattern methodology to enhance the accuracy of face recognition. We firstly propose the modified face key feature point location detection method to enhance the traditional localization algorithm to better pre-process the original face images. We put forward the grey information and the color information with combination of a composite model of local information. Then, we optimize the multi-layer network structure deep learning algorithm using the Fisher criterion as reference to adjust the network structure more accurately. Furthermore, we modify the local binary pattern texture description operator and combine it with the neural network to overcome drawbacks that deep neural network could not learn to face image and the local characteristics. Simulation results demonstrate that the proposed algorithm obtains stronger robustness and feasibility compared with the other state-of-the-art algorithms. The proposed algorithm also provides the novel paradigm for the application of deep learning in the field of face recognition which sets the milestone for further research.

Keywords: Face recognition, Deep convolution neural network, Key points detection, Local binary pattern, Feature extraction, Optimization

1. Introduction

Face recognition is one of the most important and crucial applications of image analysis and understanding which is also the human development of its biological characteristics and a great attempt. The so-called face recognition is to use features of computer analysis, video or image and extract the effective identification information, finally discriminant face the identity of the object. Role of face recognition explores various research issues and are used as practical applications include life-care systems, surveillance system, security system, face verification, patient monitoring systems and human gait recognition systems [1-5]. The face image has the very high dimension, but it may be in low-dimensional subspace or manifold. Therefore, domestic and foreign scholars put forward many subspace learning method to extract the core features of face image and uses nearest neighbor classifier and support vector machine to realize the recognition of the face image. Although these subspaces learning method has been successfully used in the field of the face recognition, however, in the face image affected by noise pollution or local with certain degree of shade, face recognition method

based on subspace learning poor performance.

Study of face recognition algorithm has for many years and the common basic algorithms of face recognition can be divided into following categories. (1) Face recognition based on subspace analysis [6-7]. Subspace method is the basic idea of high-dimensional face image feature through compressed space transformed to low dimensional subspace identification. In face recognition based on linear space it is in fact the face image expression, posture, and in the light and other complex changes in the linear simplification, it is impossible to get a face full of description. (2) The methods based on geometric features [8]. Such recognition method first will face with a geometric feature vector, according to the hierarchical clustering by using the pattern recognition thought design classifier of face recognition. In this kind of recognition based on geometric features, identify it as the matching of the feature vector based on distance measure criterion of judgment is most commonly used identification method [9-11]. (3) Face recognition based on hidden Markov model. Face recognition method based on the HMM can allow facial expression change and rotation considerably and has the high recognition rate, but at the same time, the extracted features and training but needs large computation, the model parameters and thus get less application. (4) Face recognition based on the neural network. Artificial neural network was applied to face recognition an intuitive idea is to establish the neural network as each neuron corresponds to a pixel in image.

[†] Corresponding Author: School of Information Engineering, Xijing University, China. (piaoleheng02@126.com)

* School of Information Engineering, Xijing University, China. (wjdw716@163.com)

Received: October 10, 2015; Accepted: May 16, 2016

Neural network in training is required when optimal parameters of multiple input, so the sample of the single face recognition problem is powerless [12]. (5) Face recognition based on the template matching. The basic idea of this kind of method is will face gray image as a template in database, and then by calculating the unknown samples and the normalized correlation between the known template for face recognition. Although template matching method is simple and intuitive, and the feature vector dimensions is usually face with image pixel number, therefore a large amount of calculation will be needed [13].

In vision-based technology, the depth cameras become focusing area for many researchers in different HCI fields [14-18]. Many studies have shown that, in order to be able to study high-level abstractions of complex functions to solve the target recognition, speech perception and the language understanding such as the artificial intelligence related tasks, we need to adopt deep learning technique. Deep learning architecture is composed of multilayer nonlinear computing unit, each lower output as the higher level of input that can learn from large number of input data and effective features, that many structural information contained in the input data is a kind of good method, extracted from the data can be used for classification, regression and specific problem in the information retrieval, etc. Current, some state-of-the-art research has introduces deep neural network model into the face analysis tasks. In [19], the convolutional deep belief networks are applied into face verification with the revise hierarchical representations. Study in high resolution image features, they use the convolution deep belief networks. In addition, the use of the global structure in an object class, they use local convolution restricted Boltzmann machine, a novel learning model of cyclotron, don't assume that using global stability of a character image while maintaining the scalability and robustness of small disturbance. In [20], the spatial pyramid pooling is introduced into the CNN to serve as visual recognition method. The new network structure, called SPP-net can generate a fixed-length representation regardless of image size and scale which will be beneficial for enhancing the recognition accuracy.

To enhance the traditional face recognition algorithms, we introduce the deep convolution neural network and key point detection jointed local binary pattern methodology based model in this paper. In the table one, we demonstrate the general accuracy of the recognition for traditional kinds of the approaches. The 'Accuracy 1' means the experiment without the noise influence whereas 'Accuracy 2' denotes the experiment under noisy condition. The rest of paper is organized as the follows. In the section 2 we show the core challenge of modern face recognition algorithms. In section 3, we describe the proposed algorithm for enhancing the basic accuracy of face recognition. In the section 4, we gives the simulation results and discussions on the propose model. In the section 5, we conclude our research and cast prospect for the future research.

Table 1. The Average Accuracy of the Traditional Methods

Methodology	Accuracy1	Accuracy2	Literature
Subspace Analysis	77.5%	73.0%	[6][7]
Geometric Features	81.2%	79.1%	[8]
Hidden Markov	82.9%	79.3%	[21][22]
Neural Network	93.3%	91.7%	[12]
Template Matching	85.5%	83.1%	[13]

2. Challenges for Face Recognition

Face recognition is challenging interdisciplinary frontier, at present many of the methods is still in research stage, the recognition results from practical and the larger gap. In order to conduct further in-depth study of face recognition, we summed up the field of face recognition need or subject to solve key problems and challenges in this section.

2.1 Facial expression feature recognition

In face recognition, expression is problem that not allow to ignore, because the expression changes directly affect the result of the face recognition. Using computer to realize the expression feature detection is difficult, this is because human face is a flexible body instead of rigid body which is very difficult to establish the accurate expression model. Facial expressions and the rich expression changes of subtle and complex, the abstract generalization of the characteristics of the expression change is difficult. Therefore, generally face recognition methods are based on expressionless face under the assumptions of implementation.

2.2 Light and attitude estimation problem

Light face detection and recognition is the key of a very important and difficult to solve the problem, at present a lot of recognition methods have different degree of dependence on the light conditions. Although we have put some specific solutions, on the whole, to the problem of lighting research is still relatively small as the lack of efficient and practical algorithm. Due to the diversity of facial gestures, under the natural state of human face image is not always positive, positive identification of a face image is just a kind of ideal state. In face recognition, we must consider posture change on the impact of the factor and attitude problem inevitable.

Face recognition application field expands unceasingly in recent years and many occasions require the identification based on a single face image. Overall, based on single view or the small sample attitude face recognition is current key problems needed to resolve as it directly affects the process of practical application of face recognition.

2.3 The establishment of feature fusion mechanism

Fusion technology is the use of the characteristics of the

multiple or more sensors on the complementarity and the information in time and space correlation in order to more comprehensive description of things. Integration can be on different levels in the different levels, and divided into two categories, redundant fusion and complementary fusion.

Existing face feature detection methods are effective they give face feature data from different sides, if we can really handle these characteristics for effective integration, which is an efficient way to improve the basic performance of face recognition we will achieve better accuracy. The efforts of fusion technology in the application of the face recognition with the combination of the whole and local facial feature recognition method, based on the structure of the human face recognition and fusion of different algorithms, based on the wavelet subgraph and the decision fusion of the face recognition. Information fusion theory in the application of pattern recognition such as some methods mentioned above to certain extent improves performance of face recognition, but we also need further research and improvement.

2.4 The time-varying characteristics

In face recognition, in addition to the light, gestures and expressions factors based on the time-varying factors also influence the important aspect of face recognition. Stay in practical applications, the recognition of face to face in the library faces has certain difference, but if we could change age factor applied to face recognition, the more maximum to eliminate its impact is also an important aspect of face recognition research. In general, with the facial recognition technology and method of refining, face recognition based on time-varying feature should be our practical research to an important concern.

2.5 The intensity and depth pixels problem

Image depth refers to the digits, storage for each pixel is used to measure color resolution of the image. Image depth of every pixel color images may be some color number, or to determine the gray image of each pixel gray series as the best as we can. It determines the color in the image can appear most color number or the maximum gray level in the gray image. Image content understanding depth estimation method is mainly through classifying every scene in the image block, and then for each category of scenery with respective applicable method to estimate the depth of their information [23-27].

3. Our Proposed Algorithm

This section describes the core principles of our proposed algorithm. Our model is composed of three primary parts denoted as key point detection, modified deep convolution neural network and the local binary

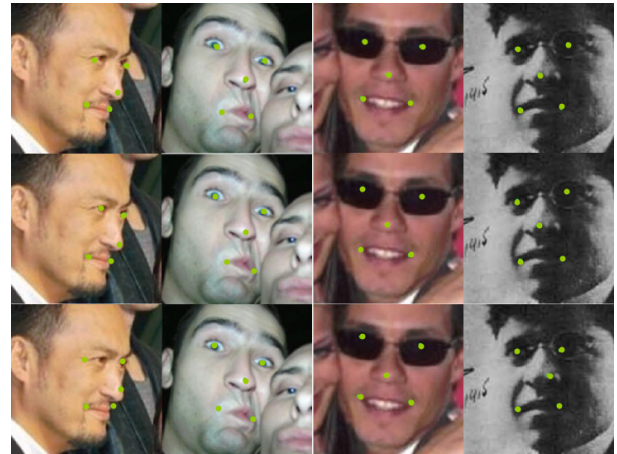


Fig. 1. The Demonstration of Key Points Detection

pattern based feature. We introduce the face point detection method to serve as the pre-processing step as the start of this section.

3.1 The key points detection

Face key feature points location detection technology is a key technology of face information pre-treatment. Feature points precisely and quickly estimate in the face recognition, gesture and facial 3D image reconstruction and so on has a very important application. In the Fig. 1, we illustrate the example of the key point detection technique.

Traditional feature point localization algorithm is based on the local gray gradient model to search the feature points, so the image gray scale is more sensitive. We have put forward the grey information and color information, the combination of a composite model of the local information. The improved method has not completely abandon the gray information, because the face skin color does not have the unique and separate from the other object color background may contain class skin color pixels, adhesion may also in the areas such as the neck and face. So, we put forward the complex local information model that can effectively and accurately distinguish the false feature points, thus improve the feature point positioning accuracy and robustness.

For training on the shape of the vector for alignment after processing, we can use the method of principal component analysis to find out statistics and regularities of variation of the shape and also achieve the representation in transform domain. An arbitrary shape vector can be approximately expressed as following expression.

$$x = \bar{x} + P_s b_s \quad (1)$$

Where, the \bar{x} represents the average shape, P_s represents the maximum eigenvalue of covariance matrix of feature vector. The vector b_s could be regarded as shape of a shape vector parameter which could be solved by the formula 2.

$$b_s = P_s^T (x - \bar{x}) \quad (2)$$

First for each training image calibration feature points along with the normal direction we use the characteristic of the calibration point as the center to take a certain number of pixels, using the grey value to form the below vector.

$$h_{ki} = \left(h_{ki0}, \dots, h_{ki(n_p-1)} \right)^T \quad K = 1, 2, \dots, n \quad (3)$$

In the formula 3, the k represents the index of the images used from the training set and i denotes the index of feature selected. Through gray vector for derivatives and after the normalized processing, we get the following equations.

$$P_{ki} = \frac{dl_{ki}}{\sum_{q=0}^{n_p-2} |l_{ki(q+1)} - l_{kiq}|}, \quad g_{ki} = \frac{dh_{ki}}{\sum_{q=0}^{n_p-2} |h_{ki(q+1)} - h_{kiq}|} \quad (4)$$

With the similar methods of extracting local gray-level information, in this paper, the partial skin based probability information is used. The difference is local skin probability model is trained in color image generated skin probability graph. We use the p_{ki} and g_{ki} to form novel feature vectors as the formula 5. The weight is adopted to balance between the numerical differential as below.

$$c_{ki} = \begin{pmatrix} g_{ki} \\ w_p p_{ki} \end{pmatrix} \quad (5)$$

The characteristic of the gray level and the chromaticity information, better than the traditional local gray features that can correctly reflect characteristic point to distribution. Therefore, the improved algorithm has the better robustness. Finally, the composite characteristics of local gradient for statistical modeling and calculating mean and covariance matrix of training image results are shown as the follows.

$$\bar{c}_i = \frac{1}{N} \sum_{k=1}^N c_{ki}, \quad \text{cov} = \frac{1}{N} \sum_{k=1}^N (c_{ki} - \bar{c}_i)(c_{ki} - \bar{c}_i)^T \quad (6)$$

It can be thought of as a normalization composite local gradient eigenvector to satisfy the Gaussian distribution as a whole. Therefore the corresponding Mahalanobis distance could be calculated through the formula 7.

$$d = (c'_i - \bar{c}_i)^T \text{cov}_i^{-1} (c'_i - \bar{c}_i) \quad (7)$$

The distance also reflects the candidate from probability of the feature points, in the process of feature point search can select probability value of the largest candidate as best match point. In the Fig. 2, we demonstrate the face key



Fig. 2. The face key points adopted by our algorithm

Table 2. The key points detection comparison

Methodology	Accuracy	Time Consuming Level
Ours	89.3%	A
[28]	85.5%	B
[29]	86.2%	A-
[30]	87.4%	A-

points adopted by our algorithm. In addition, we compare our method with the algorithms proposed in [28-30].

3.2 The modified deep convolution neural network

Deep learning concept originated in the study of artificial neural network, there is more than one hidden layer of the multilayer perceptron is a good example of deep learning model. Depth refers to the network for the neural network learning to get the number of the function of the nonlinear combination operation level. Deep learning has advantages compared with shallow learning, illustrates the necessity of introduction of deep learning.

The advantages of the deep neural network can be listed as follows. (1) In terms of information sharing, deep study the multiple levels of the extracted features can be reused in different similar tasks. (2) In terms of bionics angle, deep learning network structure is the best simulation of human brain cortex. It is the same as that of cerebral cortex and the deep learning of input data processing is layered as each layer of the neural network was used to extract the raw data characteristics of different levels. (3) As ability to express complex objective functions in network, shallow structure neural network sometimes cannot well realize the complex high-dimensional function. (4) In terms of computational complexity of the network structure, with the depth of k compact network structure to express a certain function, in the depth of less than the network structure to express the function of k, it may need to increase the number of the exponential scale factor calculation, greatly increase the complexity of the calculation [31-34].

Depth of neural network is composed of multiple single nonlinear network stacks. The encoder to provide from the input to the implicit feature space of bottom-up mapping, the decoder is to rebuild results will be implied as close as

possible to the original input for the target feature mapping to the input space. Generally, the deep neural network could be categorized as the follows. (1) Feed-back deep networks. It is composed of multiple decoder layer overlay, such as deconvolution network and hierarchical sparse coding [35]. (2) Bi-directional deep networks. By superposition of the multiple encoder and decoder layers, each layer may be a separate coding or decoding process and may also contain both encoding process also includes the decoding process [36-37]. (3) Feed-forward deep networks. It is composed of the multiple encoder layer overlay, such as the multilayer perceptron [38-39].

Convolution neural network (CNN) is the first successful training multi-layer network structure learning algorithm as based on reconstruction error of input and output energy function through the forward and back propagation network connection weights of process optimization and adjustment made to minimize the overall energy function. The forward propagation uses weights in the sharing principle to reduce the number of weight training and reduce the complexity. Backward propagation is to label value and predicted cost function is minimum error as weights of fine-tuning. The convolution process is designed to the different convolution kernels can be extracted to the different characteristics, and through the weights of sharing to reduces the computation complexity. The sampling process is equivalent to the fuzzy filter that has the effect of secondary feature extraction to reduce the dimensions and at the same time it can increase the robustness of

image feature extraction. In the Fig. 3, we illustrate the overall structure and inner organization of the modified deep convolution neural network.

Suppose that the sample sets of the obtained are shown as the following formula 8, cost of the most basic function of convolution neural network is shown as the formula 9.

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\} \quad (8)$$

$$J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|h_{w,b}(x^{(i)}) - y^{(i)}\|^2 \right) \right] \quad (9)$$

In the formula 9, b represents the bias term, W denotes the connection parameters between each unit. $h_{w,b}(x^{(i)})$ is last layer neural network output. We use gradient descent method for objective function optimization as follows.

$$b_i^{(l)} = b_i^{(l)} - \zeta \frac{\partial}{\partial b_i^{(l)}} J(W, b) \quad (10)$$

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \zeta \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b) \quad (11)$$

Where, the ζ is learning ratio. In order to make algorithm more conducive to the classification, draw lessons from the thought of the Fisher criterion, we put forward based on the distance between the class and class

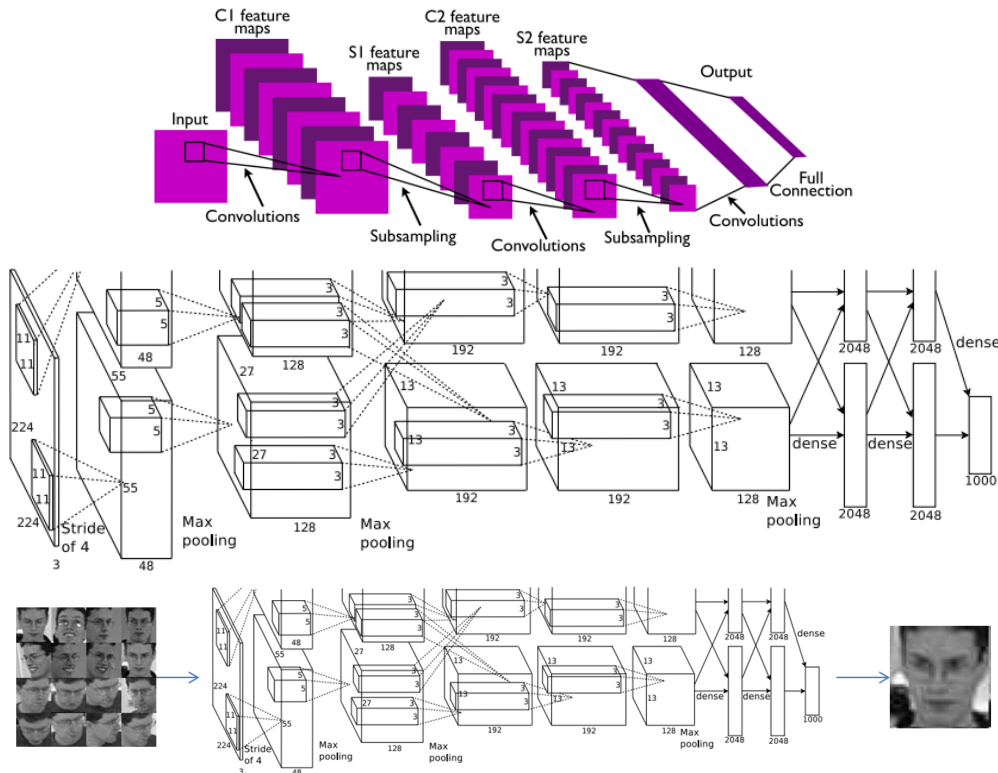


Fig. 3. The overall structure and inner organization of the modified deep convolution neural network

of energy function. In the formula 12 and 13 we demonstrate the proposed method. The \mathbb{R}_1 represents the similarity measure function in the class and the \mathbb{R}_2 denotes the similarity measure function between the different classes.

$$\mathbb{R}_1 = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n \left\| h_{w,b} \left(x^{(i,j)} \right) - M^{(i)} \right\|^2 \quad (12)$$

$$\mathbb{R}_2 = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n \left\| M^{(i)} - M^{(j)} \right\|^2 \quad (13)$$

In order to make deep learning classification algorithm characteristic of the each layer learn more conducive to join between classes within the constraints of energy function model, we revise the model into the expression 14.

$$J_{Revised} = R + \mu \mathbb{R}_1 - \gamma \mathbb{R}_2 \quad (14)$$

Where the R represents the cost function of the neural net, $J_{Revised}$ represents the overall cost function. Considering the error at the same time, make the samples in the class small spacing, spacing between the classes. In this way, when the cumulative values, it can make the weights of each layer is more advantageous to the classification of the direction of the adjustment, so that when sample size is less or training the less number of iterations get faster access to suitable for target classification.

3.3 The local binary pattern

The local binary pattern is a kind of texture description operator. The basic principle of the algorithm is selected as a threshold, the center of the image pixels by comparing the adjacent pixel gray value in the domain, through the binary code, the local texture feature of image is described using binary code. A picture of the two-dimensional face image texture feature by each pixel point in the image got by comparing with its neighborhood pixels, the coding approach is present as the formula 15.

$$f_i(x, y)_{p,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (15)$$

Where, the g_c represents the center pixel, g_p represents the uniform distribution of pixels with the center of g_c and radius of the R . After local binary pattern code of the face image, the edge of bright spots, dark spots, such as feature model is portrayed, but distribution of these patterns cannot be expressed. In order to express spatial structure features of coding figure, we cut coding diagram into small pieces. Statistical histogram of each block, gets a face of sub image histogram feature, and connects the face image feature.



Fig. 4. The local binary pattern test

$$H_{i,j} = \sum_{x,y} I\{f_i(x, y) = i\} I\{(x, y) \in R_j\} \quad (16)$$

Due to the LBP texture feature with the gray translation invariance and rotation invariance, the advantages of simple calculation, it has been successfully applied to the texture classification, face recognition, image analysis, background modeling and other fields, presents superior performance. In the Fig. 4, we show the sample test.

3.4 The final face recognition algorithm

In the pixel level face feature as the depth of the neural network input, it can study the abstract nature of implied in face images, but due to the characteristics of the input is a vector form it cannot learn local structural characteristics of face image. Deep neural network combine prior knowledge can capture facial image of local information.

The joint distribution for the current deep neural network can be expressed as the formula 17.

$$p(H, h^{(1)}, h^{(2)}, \dots, h^{(l)}) = p(H | h^{(1)}) p(h^{(2)} | h^{(1)}) \dots p(h^{(l-1)} | h^{(l)}) \quad (17)$$

When a face image affected by uniform illumination and small rotation, due to the LBP texture characteristics has the strong robust, and by the depth of neural network learning to the abstract characteristics still have invariance. The core steps of the proposed methodology are shown as follows. (1) Using bilinear interpolation will test sample and training sample and use histogram equalization and normalization pre-processing procedures. (2) Treat training samples and testing samples block to extract LBP texture characteristics of each sub-block, connect the characteristics of each sub-block form the sample LBP texture feature. (3) Use training samples of the LBP texture feature as a visual depth neural network input layer, the depth of network training, step by step to obtain the optimal network parameters. (4) When the depth after the network training, we will test sample texture features as a visual input layer, using the optimized network from down to up in learning and to extract characteristics of test sample. Back in the top of the network are classified, obtain the test sample value, and then calculate the correct recognition rate for performance evaluation.

4. Experimental Result

4.1 The data sets and environment used for testing

In the testing and verification, we adopt the four famous face recognition data sets for analysis as follows. (1) AT&T data set [40]. (2) The Georgia Tech data set [41]. (3) The Honda/UCSD data set [42]. (4) MoBo data set [43].

Before the experiment, all of the face image through eyes positioning, calibration, shearing and normalized in the end. The experimental environment configuration is six physical machines equipped with 4 TB hard disk and 6 GB of RAM, and the simulation software is installed on Windows Win7 platform and Intel core 4 quad core 2.8 GHz.

4.2 The performance under different hidden layers

Depth of neural network to the input data classification accuracy depends on the bottom, multi-level effectiveness of learning. We first discuss with different number of the



Fig. 5. The AT&T data set [40]



Fig. 6. The georgia tech data set [41]



Fig. 7. The Honda/UCSD data set [42]

hidden unit cases in this paper, algorithm performance. The table two demonstrates the result. Training time is the sum of the unsupervised total time, in the process of the training and supervised classification time refers to the test samples, the total time refers to the sum of the training time and testing time. Inspired from the table one, we could conclude that with the increase of the number of hidden units, the depth of network can better express face image feature, but the training time of the network and classification time also will increase, the amount of calculation increases gradually, and the hardware demand increases.

4.3 The performance under different sample size

In face recognition, training sample is mainly provides the prototype for the entire identification process, the more



Fig. 8. The MoBo data set [43]

Table 3. The performance evaluation under different hidden layers for proposed algorithm

Hidden Layers	Recognition Rate	Training Time/s	Classification Time/s
500	17.23%	65.13	3.23
1000	21.33%	144.64	7.11
1500	35.73%	210.94	9.65
2000	46.07%	388.26	17.12
2500	54.33%	503.21	25.22
3000	60.65%	721.36	29.87
3500	65.22%	910.12	33.96
4000	76.98%	1119.21	46.23
4500	77.12%	1302.62	50.55
5000	80.54%	1718.34	67.23
5500	83.55%	2105.32	79.33
6000	87.52%	2812.54	89.05
6500	92.21%	3509.65	95.37
7000	95.07%	4702.38	102.55

Table 4. The performance under different sample size

Algorithm	Training Sample Size				
	5	10	15	20	25
PCA	31.2%	33.5%	37.9%	41.2%	43.3%
SVM	42.9%	45.2%	51.5%	54.9%	57.2%
NN	47.3%	51.0%	54.8%	59.3%	64.8%
SRC	52.5%	57.9%	61.3%	66.2%	69.8%
LBP	64.8%	70.1%	73.8%	75.2%	80.5%
Tang [44]	80.9%	83.5%	88.7%	92.8%	95.5%
Proposed	81.7%	85.9%	91.3%	93.7%	96.8%

the more training sample is to extract the differences of the characteristics to predict the more testing phase. In the Table 4, we show the result with different sample size.

We could find when the training samples were increasing, relatively rich and category feature extracting algorithm of recognition rate soared, visible traditional algorithm under the condition of limit operations is poorer. Deep learning through unsupervised training provides good optimization after starting point for the network parameters with just a few have labeled training samples to the supervise training which can achieve good recognition effect.

4.4 The performance under different data sets

To explore the effectiveness of the algorithm in the limit conditions face recognition and universal, we use the four database data simulation respectively. The Table 5 illustrates the simulation result under evaluation parameter of overall accuracy rate. Under the optimal condition and the sufficient training time, our recognition accuracy could achieve 99.3%.

Table 5. The performance under different data sets

Algorithm	Name of the Data Set			
	AT&T	Georgia	UCSD	MoBo
PCA	40.5%	48.3%	42.5%	44.3%
SVM	51.5%	52.3%	55.8%	58.2%
NN	54.8%	56.9%	61.3%	55.2%
SRC	61.3%	63.5%	59.7%	66.3%
GMM	68.9%	69.2%	73.5%	76.6%
LBP	70.1%	74.3%	77.5%	79.9%
HMM	81.4%	88.5%	85.3%	90.2%
Tang [44]	89.7%	93.3%	91.4%	95.8%
Proposed	91.3%	94.7%	94.2%	97.4%

5. Conclusion

In this paper, a novel face recognition algorithm based on deep convolution neural network and key point detection jointed local binary pattern methodology is proposed. On the face library of experimental results show this algorithm can bottom up under the restricted condition of automatic extraction of facial image of effective characteristics. With the combination of the deep neural network and the local binary pattern, we overcome drawbacks that deep structure could not learn to face image and local characteristics. This algorithm in the pose illumination expression and sheltering comprehensive factors such as the restriction conditions has good recognition effect and optimal recognition accuracy of propose approach is 99.3%. We simulate our algorithm on the four famous face data sets compared with other methods that are state-of-the-art. The experiment results show the robustness and effectiveness of our method, our approach performs the best compared with the others simulated. In the future, we will apply our algorithm

into more data sets to optimize the universality of our algorithm.

Acknowledgements

This research was financially supported in part by the National Natural Science Foundation of China (No. 61473237) and in part by Industrial Science and Technology Research Plan in Shaanxi Province of China (No. 2016GY-141).

References

- [1] Sumathi, S., and R. Malini. "Face recognition system to enhance E Health." E-Health Networking, Digital Ecosystems and Technologies (EDT), 2010 International Conference on. Vol. 1. IEEE, 2010.
- [2] Jalal, A., and S. Kim. "Global security using human face understanding under vision ubiquitous architecture system." World Acad. Sci. Eng. Technol 13 (2006): 7-11.
- [3] Jalal, Ahmad, Shaharyar Kamal, and Daijin Kim. "A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments." Sensors 14.7 (2014): 11735-11759.
- [4] Jalal, Ahmad, and Yasir Ali Rasheed. "Collaboration achievement along with performance maintenance in video streaming." Proceedings of the IEEE Conference on Interactive Computer Aided Learning, Villach, Austria. Vol. 2628. 2007.
- [5] Jalal, A. "Security architecture for third generation (3G) using GMHS cellular network." Emerging Technologies, 2007. ICET 2007. International Conference on. IEEE, 2007.
- [6] Wang, Su-Jing, Hui-Ling Chen, Wen-Jing Yan, Yu-Hsin Chen, and Xiaolan Fu. "Face recognition and micro expression recognition based on discriminant tensor subspace analysis plus extreme learning machine." Neural processing letters 39, no. 1 (2014): 25-43.
- [7] Lu, Jiwen, Yap-Peng Tan, and Gang Wang. "Discriminative multimanifold analysis for face recognition from a single training sample per person." Pattern Analysis and Machine Intelligence, IEEE Transactions on 35, no. 1 (2013): 39-51.
- [8] Ballihi, Lahoucine, Boulbaba Ben Amor, Mohamed Daoudi, Anuj Srivastava, and Driss Aboutajdine. "Boosting 3-D-geometric features for efficient face recognition and gender classification." Information Forensics and Security, IEEE Transactions on 7, no. 6 (2012): 1766-1779.
- [9] Wang, Haoxiang, Ferdinand Shkjezi, and Ela Hoxha. "Distance metric learning for multi-camera people matching." In Advanced Computational Intelligence

- (ICACI), 2013 Sixth International Conference on, pp. 140-143. IEEE, 2013.
- [10] Ying, Yiming, and Peng Li. "Distance metric learning with eigenvalue optimization." *The Journal of Machine Learning Research* 13, no. 1 (2012): 1-26.
- [11] Ma, Lianyang, Xiaokang Yang and Dacheng Tao. "Person re-identification over camera networks using multi-task distance metric learning." *Image Processing, IEEE Transactions on* 23, no. 8 (2014): 3656-3670.
- [12] Le, Quoc V. "Building high-level features using large scale unsupervised learning." In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 8595-8598. IEEE, 2013.
- [13] Weinland, Daniel, Remi Ronfard, and Edmond Boyer. "A survey of vision-based methods for action representation, segmentation and recognition." *Computer Vision and Image Understanding* 115, no. 2 (2011): 224-241.
- [14] Kim, Joongrock, et al. "3D Multi-Spectrum Sensor System with Face Recognition." *Sensors* 13.10 (2013): 12804-12829.
- [15] Jalal, Ahmad, Jeong Tai Kim, and Tae-Seong Kim. "Development of a life logging system via depth imaging-based human activity recognition for smart homes." *Proceedings of the International Symposium on Sustainable Healthy Buildings, Seoul, Korea. 2012.*
- [16] Jalal, Ahmad, Jeong Tai Kim, and Tae-Seong Kim. "Development of a life logging system via depth imaging-based human activity recognition for smart homes." *Proceedings of the International Symposium on Sustainable Healthy Buildings, Seoul, Korea. 2012.*
- [17] Jalal, A. S., and Shaharyar Kamal. "Real-time life logging via a depth silhouette-based human activity recognition system for smart home services." *Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on. IEEE, 2014.*
- [18] Jalal, Ahmad, Shaharyar Kamal, and Daijin Kim. "Shape and motion features approach for activity tracking and recognition from kinect video camera." *Advanced Information Networking and Applications Workshops (WAINA), 2015 IEEE 29th International Conference on. IEEE, 2015.*
- [19] Huang, Gary B., Honglak Lee, and Erik Learned-Miller. "Learning hierarchical representations for face verification with convolutional deep belief networks." In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2518-2525. IEEE, 2012.
- [20] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Spatial pyramid pooling in deep convolutional networks for visual recognition." In *Computer Vision–ECCV 2014*, pp. 346-361. Springer International Publishing, 2014.
- [21] Abdel-Hamid, Ossama, Abdel-rahman Mohamed, Hui Jiang, and Gerald Penn. "Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition." In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 4277-4280. IEEE, 2012.
- [22] Morwal, Sudha, Nusrat Jahan, and Deepti Chopra. "Named entity recognition using hidden Markov model (HMM)." *International Journal on Natural Language Computing (IJNLC)* 1, no. 4 (2012).
- [23] Min, Rui, et al. "Real-time 3D face identification from a depth camera." *Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012.*
- [24] Jalal, Ahmad, Jeong Tai Kim, and Tae-Seong Kim. "Human activity recognition using the labeled depth body parts information of depth silhouettes." *Proceedings of the 6th International Symposium on Sustainable Healthy Buildings, Seoul, Korea. 2012.*
- [25] Jalal, Ahmad, Md Zia Uddin, and T-S. Kim. "Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home." *Consumer Electronics, IEEE Transactions on* 58.3 (2012): 863-871.
- [26] Cai, Qin, et al. "3d deformable face tracking with a commodity depth camera." *Computer Vision–ECCV 2010. Springer Berlin Heidelberg, 2010. 229-242.*
- [27] Jalal, A. S., Yeonho Kim, and Daijin Kim. "Ridge body parts features for human pose estimation and recognition from RGB-D video data." *Computing, Communication and Networking Technologies (ICCCNT), 2014 International Conference on. IEEE, 2014.*
- [28] Jun, Bongjin, Inho Choi, and Daijin Kim. "Local transform features and hybridization for accurate face and human detection." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35.6 (2013): 1423-1436.
- [29] Jalal, Ahmad, and Amir Shahzad. "Multiple facial feature detection using vertex-modeling structure." *Proceedings of the IEEE Conference on Interactive Computer Aided Learning, Villach, Austria. Vol. 2628. 2007.*
- [30] Erdogmus, Nesli, and Sébastien Marcel. "Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect." *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on. IEEE, 2013.*
- [31] Li, Kehuang, Zhen Huang, You-Chi Cheng, and Chin-Hui Lee. "A maximal figure-of-merit learning approach to maximizing mean average precision with deep neural network based classifiers." In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 4503-4507. IEEE.
- [32] Huang, Jui-Ting, Jinyu Li, Dong Yu, Li Deng, and Yifan Gong. "Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers." In *Acoustics, Speech and Signal*

- Processing (ICASSP), 2013 IEEE International Conference on, pp. 7304-7308. IEEE, 2013.
- [33] Sainath, Tara N., Brian Kingsbury, Abdel-rahman Mohamed, and Bhuvana Ramabhadran. "Learning filter banks within a deep neural network framework." In Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on, pp. 297-302. IEEE.
- [34] Saon, George, Hagen Soltau, David Nahamoo, and Michael Picheny. "Speaker adaptation of neural network acoustic models using i-vectors." In Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on, pp. 55-59. IEEE.
- [35] Ray, Jessica, Brian Thompson, and Wade Shen. "Comparing a high and low-level deep neural network implementation for automatic speech recognition." High Performance Technical Computing in Dynamic Languages (HPTCDL), 2014 First Workshop for. IEEE, 2014.
- [36] Gerstenberger, M. C., A. Christophersen, R. Buxton, and A. Nicol. "Bi-directional risk assessment in carbon capture and storage with Bayesian Networks." International Journal of Greenhouse Gas Control 35 (2015): 150-159.
- [37] Llorente, Roberto, Maria Morant, Eloy Pellicer, Milan Herman, Zsolt Nagy, Javier Herrera, Jose Correcher et al. "Radio-over-fiber quintuple-play service provision for deep fiber-to-the-home passive networks." In Communications Workshops (ICC), 2013 IEEE International Conference on, pp. 868-872. IEEE, 2013.
- [38] Sivaram, Garimella SVS, and Hynek Hermansky. "Sparse multilayer perceptron for phoneme recognition." Audio, Speech, and Language Processing, IEEE Transactions on 20, no. 1 (2012): 23-29.
- [39] Panchal, Gaurang, Amit Ganatra, Y. P. Kosta, and Devyani Panchal. "Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers." International Journal of Computer Theory and Engineering 3, no. 2 (2011): 332-337.
- [40] Oh, Sung-Kwun, Sung-Hoon Yoo, and Witold Pedrycz. "Design of face recognition algorithm using PCA-LDA combined for hybrid data pre-processing and polynomial-based RBF neural networks: Design and its application." Expert Systems with Applications 40.5 (2013): 1451-1466.
- [41] Wang, Su-Jing, Jian Yang, Ming-Fang Sun, Xu-Jun Peng, Ming-Ming Sun, and Chun-Guang Zhou. "Sparse tensor discriminant color space for face verification." Neural Networks and Learning Systems, IEEE Transactions on 23, no. 6 (2012): 876-888.
- [42] Hu, Yiqun, Ajmal S. Mian, and Robyn Owens. "Face recognition using sparse approximated nearest points between image sets." Pattern Analysis and Machine Intelligence, IEEE Transactions on 34, no. 10 (2012): 1992-2004.
- [43] Lu, Jiwen, Gang Wang, and Philippe Moulin. "Image set classification using holistic multiple order statistics features and localized multi-kernel metric learning." In Computer Vision (ICCV), 2013 IEEE International Conference on, pp. 329-336. IEEE, 2013.
- [44] Cao, Zhimin, et al. "Face recognition with learning-based descriptor." Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010.



Wen-zhun Huang was born in Hubei province, China. He received the B.S. and M.S. degrees in Communication and information system from Air Force Engineering University, China, in 1997. He received the Ph.D. degree in Information and communication engineering from Northwest Polytechnic University, China, in 2010. He is an associate professor in Xijing University. His research interests include information processing, anti-jamming technology, wireless communication system and IoT technology.



Shan-wen Zhang was born in Shaanxi province, China. He received the B.S. degree in mathematics from Northwest University, China, in 1988. He received the M.S. degree in applied mathematics from Northwest Polytechnic University, China, in 1995. He received the Ph.D. degree in electromagnetic field and microwave from Air Force Engineering University, China, in 2001. He is a professor in Xijing University. His research interests include wavelet transforms, rough sets, genetic algorithm, pattern recognition and manifold learning.