Journal of Korea Multimedia Society Vol. 21, No. 4, April 2018(pp. 471-477)
https://doi.org/10.9717/kmms.2018.21.4.471

Human Face Recognition used Improved Back-Propagation (BP) Neural Network

Ru-Yang Zhang[†], Eung-Joo Lee[††]

ABSTRACT

As an important key technology using on electronic devices, face recognition has become one of the hottest technology recently. The traditional BP Neural network has a strong ability of self-learning, adaptive and powerful non-linear mapping but it also has disadvantages such as slow convergence speed, easy to be traversed in the training process and easy to fall into local minimum points. So we come up with an algorithm based on BP neural network but also combined with the PCA algorithm and other methods such as the elastic gradient descent method which can improve the original network to try to improve the whole recognition efficiency and has the advantages of both PCA algorithm and BP neural network.

Key words: Face Recognition; PCA Algorithm; BP Neural Network; Additional Momentum; Elastic Gradient Descent Method

1. INTRODUCTION

In recent years face recognition has gained a lot of attention. In this paper, we use BP neural network, BP, also known as back-propagation, is a multilayer feed-forward neural network using error back propagation. It has advantages such as self-learning, adaptive and powerful non-linear mapping capabilities[1]. But it also has serious disadvantages like slow convergence speed and easy to fall into local minimum points. So in this paper, we try to improve that with a PCA face recognition algorithm based on improved BP network. We used PCA algorithm to extract the main features of face images and other two methods to try to avoid those disadvantages. And finally trying to improve recognition efficiency and speed[2].

2. PCA AND BP ALGORITHM

2.1 PCA feature extraction

PCA is a way to simplify the data structure by reducing the data dimension. Its essence is to high-dimensional space in the data through the orthogonal transformation projection to low-dimensional space, so as to achieve the main features of the image data extraction. The principal component analysis method is used to solve a feature subspace, which will be used to train and test the image to be projected in the feature subspace. The main feature (ie, projection coefficient) for distinguishing different faces can be obtained as the feature vector. The main features of the extraction can be identified and classified by a classifier[3].

2.2 The fundamental of BP neural network

BP neural network is a multilayer feed-forward

※ Corresponding Author : Eung-Joo Lee, Address: 428, Sinseon-ro, Nam-gu, Busan, Korea, TEL : +82-51-629-1143, E-mail : ejlee@tu.ac.kr
Receipt date : Mar. 12, 2018, Revision date : Apr. 2, 2018
Approval date : Apr. 6, 2018

[†] Dept. of Information & Communication Eng., Graduate School, Tongmyong University
(E-mail : dlzry@naver.com)
[††] Dept. of Information & Communication Eng., Tongmyong University

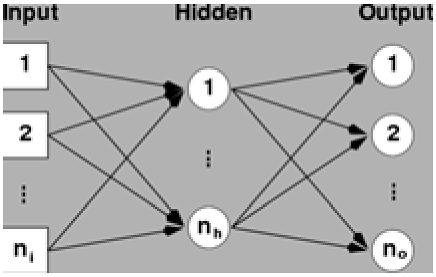

Fig. 1. The Three Levels of BP Neural Network.

neural network with error back propagation. Robert Hecht-Nielson proves that a three-layer BP neural network with a hidden layer can effectively approximate any continuous function. Among them, the three-tier BP network includes input layer, hidden layer, output layer[4]. The basic structure is shown in Fig. 1.

The basic idea of BP neural network is to use the forward propagation of the signal and the reverse propagation of the error. In forward propagation, the incoming data from the input layer is processed at each hidden layer and finally reaches the output layer. If the output is not expected, the error of the output layer is propagated backward as the adjustment signal, and the weights and thresholds are adjusted according to the prediction error, and finally, the output of the network can accept the accuracy. BP neural network has self-learning, adaptive and powerful non-linear mapping capabilities[5].

3. PCA FACE RECOGNITION BASED ON BP NEURAL NETWORK

In the traditional BP network, the gradient descent algorithm is used to overcome the problem of slow convergence, training process and easy to fall into the local minima in the learning process. In view of the above problems, this paper proposes a weight adjustment method based on the elastic gradient method. The BP neural network is improved by this weight adjustment method to speed up the convergence speed and learning precision of the network[6].

The traditional BP neural network in the adjustment of the weight of the process did not consider the previous weight adjustment, which often makes the learning process prone to shock, not easy to obtain the global optimal solution[7]. In order to reduce the shock problem in the training process, we can introduce a momentum item to reflect the accumulation of past experience. The weight adjustment of the BP network is related not only to the gradient of the error term but also to the previous weight adjustment. At this time the weight adjustment formula is:

$$\triangle W(t+1) = (1-mc)\eta\delta x + mc\triangle W(t) \tag{1}$$

Where $\triangle W(t)$ is the change of the weight, t is the number of training times, the momentum coefficient is expressed by mc, η is the learning rate, δ is the error term, x is the input quantity, δx essentially reflects the gradient of the error term to the weight, it can be seen that the additional momentum method takes into account the effect of previous empirical accumulation on the weight adjustment, where the momentum coefficient $mc \in (0,1)$. By introducing a momentum term. the previous adjustment experience can be added to the adjustment process of the weight, thus reducing the sensitivity of the network to the local minimum point. When the error surface suddenly changes, it is easier to skip the local pole smaller.

We can also use the Elastic gradient descent method[8], the adjustment of the weight is related to the update value $\triangle t$, and the gradient only affects the direction of the weight adjustment. This can effectively avoid the gradient size caused by the slow convergence of the network. Where the relationship between the size of the weight adjustment and the updated value is:

$$\triangle W(t) = \begin{cases} -\triangle t & \dfrac{\vartheta E(t)}{\vartheta W} > 0 \\ +\triangle t & \dfrac{\vartheta E(t)}{\vartheta W} < 0 \\ 0 & \dfrac{\vartheta E(t)}{\vartheta W} = 0 \end{cases} \tag{2}$$

Then the corresponding t+1 time weight adjustment formula is:

$$W(t+1) = W(t) + \triangle W(t) \tag{3}$$

The adjustment rule for the update value $\triangle t$ is:

$$\triangle t = \begin{cases} \alpha \times \triangle(t-1) \dfrac{\vartheta E(t)}{\vartheta W} \times \dfrac{\vartheta E(t-1)}{\vartheta W} > 0 \\ \beta \times \triangle(t-1) \dfrac{\vartheta E(t)}{\vartheta W} \times \dfrac{\vartheta E(t-1)}{\vartheta W} < 0 \\ \triangle(t-1) \dfrac{\vartheta E(t)}{\vartheta W} \times \dfrac{\vartheta E(t-1)}{\vartheta W} = 0 \end{cases} \tag{4}$$

And $0 < \beta < 1 < \alpha$.

The adjustment of the updated value is related to the direction of the two successive gradients. If the adjustment direction is the same for two consecutive times, the update value is incremented. If the adjustment direction is reversed for the second time, the update value is reduced and the other update value remains unchanged. By introducing the elastic gradient descent method to overcome the gradient size of the adverse effects on the network, so that the network can quickly converge.

Although those two methods can improve the BP neural network, they both have their own disadvantages. Introduce a momentum item: Training time is too long; The Elastic gradient descent method: Easy to fall into the global optimal solution[9]. So we combined those two methods together, the adjustment formula of the weight of moment t+1 is:

$$W(t+1) = \begin{cases} W(t) - sign(\dfrac{\vartheta E(t)}{\vartheta W}) \triangle t & \dfrac{\vartheta E(t)}{\vartheta W} \neq 0 \\ 0 & \dfrac{\vartheta E(t)}{\vartheta W} = 0 \end{cases} \tag{5}$$

And the adjustment formula of update value is:

$$\triangle t = \begin{cases} \alpha \times (1-mc) \times \triangle(t+1) + mc \times \triangle(t+1) \\ \beta \times (1-a) \times \triangle(t-1) \times sign(\dfrac{\vartheta E(t)}{\vartheta W}) + a \times \triangle(t-1) \end{cases} \tag{6}$$

4. IMPROVED PCA FACE RECOGNITION ALGORITHM FOR BP NEURAL NETWORK

Firstly, the principal component analysis is used to map the image data in the high-dimensional space to the low-dimensional space, and the main features in the face map image are extracted. Then, the weighting method combined with the additional momentum and the elastic gradient descent method The BP network is improved to train the main features of the face database image. Finally, the test image is used to test the trained network and obtain the corresponding recognition result. The specific steps of the model are as follows:

Step 1: Read the images in the face database, use M images for training and M' images for testing. An image with a resolution of $m \times n$ is connected to each column to form a column vector of $N = m \times n$ dimensions. The image used for training in the entire face database can be stored in a matrix $R^{N \times M'}$. Similarly, the image used for testing in the entire face database is stored in a matrix $S^{N \times M'}$. $(J = 1,2,...,M')$ The image X_j' is stored in the j-th column of the matrix $S^{N \times M'}$. The average face for training the image can be obtained as:

$$\overline{X} = \frac{1}{M} \sum_{i=1}^{M} x_i \tag{7}$$

Step 2: A covariance matrix is constructed according to the difference between the image used for training and the average face:

$$C = \frac{1}{M} \sum_{i=1}^{M} (x_i - \overline{x})(x_i - \overline{x})^T AA^T \tag{8}$$

Step 3:Solving a Feature Subspace of Image Projection. The covariance matrix C is a $N \times N$ matrix. Since N is a large number, it is difficult to directly solve the eigenvalues and eigenvectors of the covariance matrix C. In order to simplify the computational complexity, the method of singular value decomposition Solve the EAT and the eigenvector of the matrix AA^T to solve the AA^T indirectly. The characteristic values of A^T which are λare sorted in descending order. The ratio of the r eigenvalues and the total eigenvalues is the eigenvalue contribution rate ϕ, ie, $\phi = \sum_{i=1}^{r} \lambda_i / \sum_{i=1}^{M} \lambda_i$.

The eigenvalue of eigenvalue contribution rate $\phi \geq P\%$ is chosen and the corresponding eigenvector is constructed to construct a characteristic subspace. $(I=1,2,...,r)$ is the r eigenvalue of the matrix A^T, and $v_i\,(i=1,2,...,r)$ is the eigenvalue corresponding to the former r eigenvalues, Decomposition Theorem know AA^T is:

$$\overline{X}=\frac{1}{\sqrt{\lambda_i}}Av_i \qquad i=1,2,...,r \tag{9}$$

And then we can get a Feature subspace $U=(u_1,u_2,...,u_r)$

Step 4: The difference between the image used for the training and the average face is projected to U^t to obtain the corresponding projection coefficient:

$$Y=U^T(x_i-\overline{x}) \qquad i=1,2,...,M \tag{10}$$

Step 5: The difference between the image used for the test and the average face is projected to the U^T to obtain the corresponding projection coefficient:

$$U^TY'=U^T(x'^j-\overline{x}) \qquad j=1,2,...,M \tag{11}$$

Step 6: The projection coefficients Y and Y' are normalized, and the normalized Y is used as the input of the network to train the BP neural network. The number of neurons in the BP neural network input layer is determined by the dimension of the projection coefficient Y. The number of neurons in the hidden layer needs to be determined based on experience and multiple tests, and the output neurons are determined by the number of face categories in the face database. The weight of the network is used in the text of the weight adjustment method.

Step 7: According to the trained network, the number of training sets of images N and the number N' of test sets is correctly classified according to the projection coefficients of the test set face image in the characteristic subspace.

Step 8: Finally, the image recognition rate $(N/M)\times 100\%$ for training is obtained by calculating the ratio of the correct number of training for the training image and the total number of training images and the ratio of the correct number of test images to the total number of test images. The image recognition rate $(N'/M')\times 100\%$ of the test.

5. EXPERIMENT RESULTS

This paper chooses ORL face database for face recognition experiment[10]. The database has a total of 40people face images, each with 10 of 112×92 images, a total of 400 images. People's facial expressions and details have a certain change, such as laughing or not laughing, eyes open or closed, wearing glasses or not wearing glasses; the other face of the posture also has a great change, the depth of rotation and plane rotation up to 20 degrees, face size also has the most 10% change. ORL face database part of the image shown in Fig. 2.

From the image resolution of 112×92 we can see, $m=112, n=92, N=10304$. In each person's face image randomly selected 8 for training, the other two for testing. Then $M=320$. Based on

$$\overline{X}=\frac{1}{M}\sum_{i=1}^{M}x_i \tag{12}$$

Fig. 2. Part of the image in ORL Database.

The image in the library uses the principal component analysis method to extract the main features and finds a set of feature face space. The eigenvalues are arranged in descending order, the average face of the training set image and the eigenvalues of the first, second, 20th, 60th, 150th, 200th, 240th, 280th, 310th of the eigenvalue of the image corresponds to the eigen face image[11].

From the above figure, we can see that the larger eigenvalues reflect the general contours of the images and the smaller eigenvalues, and more reflect the details of the image.

A feature subspace is constructed by using the eigenvalues corresponding to the r eigenvalues, and then the images for training and the images of the test are projected to the space respectively to obtain the corresponding projection coefficients. The projection of the training image is normalized to $[-1,1]$ as the input of the improved BP neural network, the training of the neural network is carried out, and the face image on the test set is normalized and then identified. In this paper, the number of neurons in the hidden layer is 100, the number of neurons in the output layer is 40, the learning rate is 0.03, the momentum coefficient $mc=0.5$, and a is satisfying $a \in (1,1/3)$ Random number, training 1 500 times, coefficient $\alpha=1.2$, coefficient $\beta=0.5$, Fig. 3 shows the use of traditional BP neural network as a classifier and improved BP neural network classifier to select the number of eigenvectors r and test set image recognition rate [12] (Note: The recognition rate mentioned in this article is to run the program 10times the average recognition rate).

It can be seen from Table 1 that the algorithm of improving BP network proposed in this paper has higher recognition rate than the traditional algorithm, and the recognition rate of the image increases with the number of eigenvectors. When the number of eigenvectors is 40, the recognition rate of the improved algorithm is as high as 97.1%. But after that, the recognition rate of the image has a small range of fluctuation with the increase of the number of eigenvectors. This shows that some of the features introduced may not be able to distinguish the main features of the image when the number of feature vectors is increased to a certain extent. It also shows that the recognition rate of the image is not proportional to the number of eigenvectors. Find the number of feature vectors with the highest recognition rate. In the process of running the program, the average running time of the traditional algorithm is 23.4s, and the average running time of the recognition algorithm is 1.3s. The results show that the proposed algorithm has faster convergence rate and higher recognition rate than traditional algorithms.

In addition, the $k(k=1,2,....,9)$ images of each person in the face database are randomly selected for training, and the other 10-k images are used

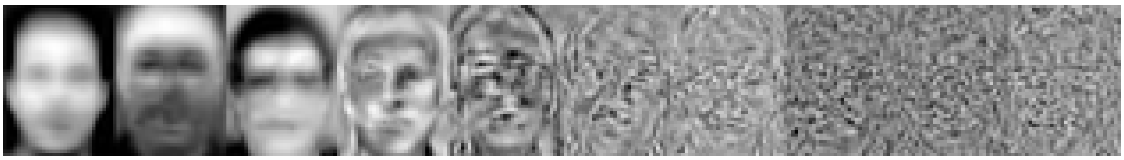

Fig. 3. The average face.

Table 1. The relationship between the number of training images and the recognition rate

Test image recognition rate (%)	The number of feature vectors									
	5	10	15	20	25	30	35	40	45	50
Traditional Algorithm	57.2	74.1	81.3	83.4	86.2	87.6	89.1	90.9	91.2	92.8
Proposed Algorithm	78.5	89.3	92.1	92.8	93.2	95.7	96.5	97.1	97.0	97.1

Table 2. The relationship between the number of training images and the recognition rate

Test image recognition rate (%)	The number of training images								
	1	2	3	4	5	6	7	8	9
Traditional Algorithm	37.7	58.6	70.1	80.3	87.2	90.6	92.2	94.9	95.8
Proposed Algorithm	60.6	78.3	87.1	90.5	92.6	93.7	94.3	96.1	97.7

for testing, and eigenvalues with eigenvalues of greater than 85% and the corresponding eigenvector. The relationship between the number of training images used in each class and the recognition rate on the test image is obtained as shown in Table 2.

It can be seen from Table 2 that the improved algorithm proposed in this paper has a higher recognition rate than the traditional algorithm, and the number of training images has a great influence on the recognition rate of the test image. As the number of training images in each category increases, the recognition rate on the test image increases. When the number of images used for training in each face database is 1, and the number of test images is 9, the recognition rate of the improved algorithm proposed in this paper is 60.6% on the test image. The classification ability of the network is increasing with the increase of the number of training images. In particular, when the number of images used for training in each face database is 9, and the number of test images is 1, the recognition rate on the test image is 97.7%

6. CONCLUSION

Through the experiment results, we can see that the improved algorithm proposed in this paper can be used to analyze the images in the ORL face database and also can work faster and more efficient than the traditional algorithm. And using the improved algorithm that proposed in this paper when we run the face recognition equipment, we can get what we want faster and more accurate. In the future research, we can try to fix the deficiency of the existing algorithm and find more ways to improve the speed and the recognition rate based on the existing methods.

REFERENCE

[1] H.H. Nam, B.J. Kang, and K.H. Park, "Comparison of Computer and Human Face Recognition According to Facial Components," *Journal of Korea Multimedia Society*, Vol. 37, No. 21, pp. 40-50, 2012.

[2] H.J. Moon, and S.H. Kim, "Comparison of Computer and Human Face Recognition According to Facial Components," *Journal of Korea Multimedia Society*, Vol. 6, No. 2, pp. 247-258, 2013.

[3] W.O. Lee, Y.H. Park, E.C. Lee, H.K. Lee, and K.R. Park, "Tracking and Face Recognition of Multiple People Based on GMM, LKT and PCA," *Journal of Korea Multimedia Society*, Vol. 15, No. 4, pp. 449-471, 2012.

[4] Z. Li, and X. Fu, "Research on Performance of Three Improved BP Algorithms Based on PCA," *Journal of Computer Engineering*, Vol. 37, No. 21, pp. 108-110, 2011.

[5] R. Hecht-Nielsen, "Theory of the Back Propagation Neural Network," *Journal of the International Joint Conference on Neural Networks*, Vol. 1, No. 1, pp. 455, 1988.

[6] M. Kirby, and L. Sirovich, "Application of the Karhunen-loeve Procedure for the Characterization of Human Faces," *Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, pp. 103-108, 2002.

[7] M. Riedmiller, and H. Braun, "A Direct Adaptive Method for Faster Back Propagation Learning: The RPROP Algorithm," *Proceedings of the IEEE International Conference on*

Neural Networks, Vol 1, pp. 586-591, 1993.

[8] Z.D. Long, G. Wen, and D.B. Zhao, et al., "Faces Recognition Based on Singular Value Decomposition and Discriminant KL Projection," *Journal of Software*, Vol. 14, No. 4, pp. 783-788, 2003.

[9] Y.S. Xu, J.H. Gu, Z. Tao, D. Wu, and M.C. Zhu, "Handwritten Character Recognition Based on Improved BP Neuralnetwork," *Journal of Communication Technology*, Vol. 5, No. 44, pp. 106-109, 2011.

[10] C.J. Zhao, "Research on Expression Recognition Based on Optimized BP Neural Network," *Proceedings of International Conference on Industrial Engineering and Engineering Management*, pp. 1803-1806, 2009.

[11] L. Yan, "Fusion Method of PCA and BP Neural Network for Face Recognition," *Proceedings of the International Conference on Computer Science and Service System*, pp. 3256-3259, 2011.

[12] B. Messaoud, M. Lamia, H. Farid, and M. Cheriet, et al., "A Guessoum. Face Recognition Based on 2DPCA, DIAPCA and DIA2 DPCA in DCT Domain," *Proceedings of the International Multi-conference on Systems*, Signals and Devices pp. 1-6, 2008.

Ruyang Zhang

He received his B.S. at Dalian Polytechnic University in China (2012-2016). Currently he is studying in the Department of Information and Communication Engineering in Tongmyong University, Korea for his Master degree. His main research areas are image processing and face recognition.

Eung-Joo Lee

He received his B.S., M.S.,Ph.D. in Electronic Engineering from Kyungpook National University, Korea, in 1990,1992, and Aug. 1996, respectively. Since 1997, he has been with the department of Information & Communication Engineering, Tongmyong University, Korea, where he is currently a professor. From 2000 to July 2002, he was a president of Digital Net Bank Inc. From 2005 to July 2006, he was a visiting professor of the Department of Computer and Information Engineering, Dalian Polytechnic University, China. His main research interests include biometrics, image processing and computer vision.