# Illumination Invariant Face Recognition Based on Neural Network Ensemble

Wu-Jun Li[1], Chong-Jun Wang[1], Dian-Xiang Xu[2], and Shi-Fu Chen[1]
[1]*National Laboratory for Novel Software Technology*
*Nanjing University, Nanjing 210093, P.R.China*
*Email: {liwujun, chjwang}@ai.nju.edu.cn*
[2]*Department of Computer Science*
*North Dakota State University, Fargo, ND58105, USA*
*Email: Dianxiang.xu@ndsu.nodak.edu*

## Abstract

*An illumination invariant face recognition method based on neural network ensemble architecture is proposed. Given a face image with an arbitrary illumination direction, it can complete recognition in a uniform way with high performance without knowing or estimating the illumination direction. Experimental result shows that the recognition ratio of the ensemble architecture is higher than the conventional approach that uses a single neural network to recognize faces of a specific illumination direction.*

## 1. Introduction

Face recognition is to identify or verify one or more persons in the given still or video images of a scene using a stored database of faces [1]. Existing approaches to face recognition can be classified into two categories [2]: geometric feature-based and appearance-based. The geometric feature-based methods, such as elastic bunch graph matching [3] and active appearance model [4], make use of the geometrical parameters that measure the facial parts; whereas the appearance-based methods use the intensity or intensity-derived parameters. As a representative of this approach, the eigenface approach [5] has become the benchmark of face recognition techniques.

The FERET evaluation [6], however, shows that the performance of a face recognition system may decrease seriously with the change of illumination condition or pose. To recognize faces under variable pose, Pentland et al [7] have proposed a multiple eigenspaces method, which builds view-specific eigenspaces. For recognition of human faces with any view in a certain viewing angle range, Huang et al [8] have introduced neural network ensemble to multiple eigenspaces.

Still, varying illumination is a challenge to the face recognition community. To address the variable illumination issue, this paper generalizes the neural network ensemble architecture [8]. Specifically, we build multiple eigenspaces in terms of illumination directions and train illumination direction-specific neural networks on the feature coefficients projected in the corresponding eigenspaces. All illumination direction-specific neural networks are then combined by a neural network ensemble module. At the test phase, given one input image with an arbitrary illumination direction, the ensemble architecture can complete recognition in a uniform way that we feed the input image into these different channels corresponding to different illumination directions and obtain a final decision from the ensemble module. Experimental result shows that the recognition accuracy of our method is higher than the conventional approach that uses a single neural network to recognize faces of a specific illumination direction.

## 2. Multiple Illumination Eigenspaces

Assume that the face images in a training set, say X, can be divided into a number of subsets, say $\{X1,X2,...,XI\}$ , according to the illumination directions when images are imaged, i.e. $X=\{X1,X2,...,XI\}$, where $Xi=\{x_{i1},x_{i2},...,x_{in_i}\}$ , I is the number of different illumination directions, $x_{ij}$ denotes the *jth* image corresponding to the *ith* illumination direction, and $n_i$ is the number of face images corresponding to the *ith* illumination direction. We can build I eigenspaces [5], each describing a particular region of the facespace that is corresponding

Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004)
1082-3409/04 $20.00 © 2004 **IEEE**

to a specific imaging illumination direction. Thus, we can get the corresponding eigenspace set, say $S=\{S_1,S_2,...,S_I\}$, where $S_i$ stands for an eigenspace corresponding to the *ith* illumination direction. If I > 1, we have multiple illumination eigenspaces.

## 3. Neural Network Ensemble for Illumination Invariant Face Recognition

We extract the feature coefficients of each image in the corresponding illumination eigenspace. Then illumination direction-specific backpropagation (BP) neural networks with one hidden layer are trained on the extracted feature coefficients. Decision making is done by an ensemble module that combines the illumination direction-specific neural networks (we call it the first layer neural networks). The ensemble architecture is shown in Figure 1. The first layer neural networks are trained on the feature coefficients of the corresponding images extracted in the corresponding eigenspaces. At the test phase, given one input image, we feed it into these different channels and obtain a final decision from the ensemble module.

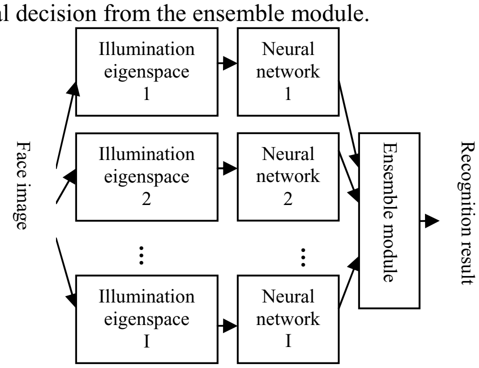


**Figure 1: Neural Network Ensemble Architecture**

Considering that we use multiple eigenspaces to extract the feature coefficients of faces, it is natural to train a neural network for each eigenspace, and then combine the decisions of all individual neural networks. If we only use one neural network, there would be much collision of data. For example, when the image of object *A* with illumination direction *I* is projected in eigenspace *Si*, the feature vector would probably be *V*, and when the image of object *B* with illumination direction *J* is projected in eigenspace *Sj*, the feature vector would probably be *V* too. In this case, the neural network would not converge because of the existence of data collision.

### 3.1. Selection of Ensemble Method

There are many methods for combining the results of individual neural networks. The most prevailing approaches include plurality voting or majority voting for classification tasks [9] and simple averaging [10] or weighted averaging [11] for regression tasks. But none of these methods are suitable for the varying illumination problem of face recognition. Voting is effective when all of the individual neural networks are used to classify the same problem. In our case, each neural network only deals with a specific sub-problem of the whole problem. As shown in Figure 2, voting is not effective: the expected result is *"A"*, but voting would result in "*reject*".

Simple averaging and weighted averaging are not suitable for our problem either. They can only combine multiple individual neural networks in a linear fashion. The correlation among all of the multiple illumination eigenspaces, however, is obviously nonlinear. Therefore we use a backpropagation (BP) neural network with one hidden layer to realize the ensemble module in Figure 1.

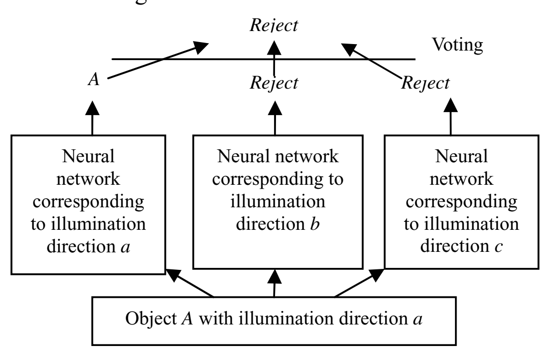


**Figure 2: Result of Voting**

### 3.2. Architecture of the Neural Network of Ensemble Module

We call the neural network of the ensemble module the second layer neural network. Its architecture is shown in Figure 3. The input vector of the second layer is the cascade of all the output vectors of the first layer. Suppose *n* eigenspaces are used, and the number of output units of each neural network in the first layer is *k* (each output unit stand for an object to be recognized). Thus, there are *q* ($q=n*k$) input units for the second layer. If these *q* input units are arranged into an $n*k$ array, each column of the array will stand for a specific object to recognize, whereas each row of the array will stand for a specific illumination direction. The second layer is used as a classifier that has *k* output units, each of which stands for an object. At the training phase, we force the output unit corresponding to the correct person to be “1” and all the other units’ values to be “0”.

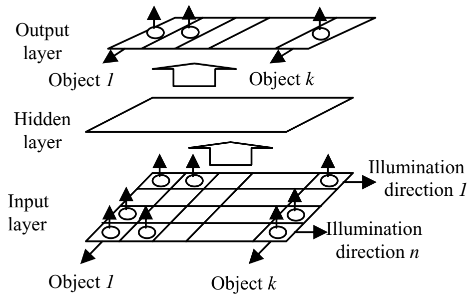


**Figure 3: The Second-layer Neural Network**

We cut off the thresholding part of the output units in the first layer neural networks, as in [8]. That is, we train them as classifiers and use them as regression estimators.

## 4. Experiment

While a number of benchmark face bases for face recognition, such as FERET database [6], are available, none of them are suitable for face recognition methods that are based on multiple illumination eigenspaces and neural networks, because it need multiple images of each subject with each specific illumination direction to build multiple eigenspaces and train neural networks. To evaluate our approach, we have built our own face base.

### 4.1. Data Acquisition

In our experiment, we have collected indoor face images using a Logitech QuickCam® Pro 3000 camera. Considering that illumination in the real world usually consists of an ambient light with perhaps one or two point sources, we kept the room lights on when capturing images in order to make the images look natural. Room lights were not very bright to irradiate a subject’s face uniformly. When all imaging conditions except illumination remained unchanged, face images with illumination direction from right can be obtained from images with illumination direction from left by reflection transformation. Therefore only those images with illumination from left are considered in this paper. We call the front of a face 90’ direction, the left of a face 0’ direction. Accordingly, there are 30’ and 60’ directions in the left front of a face. To collect face images with different illumination directions, we let subjects sit in front of the camera, using a desk lamp with a meter far from the person as light source. Putting the light source in the four directions mentioned before, we collected face images for everybody sitting in front of the camera. During the process of collecting images for a subject, the subject kept almost facing the camera and made various expressions. For those persons who wore glasses, we collected their face images with glasses on and off respectively. We have collected face images of 27 subjects consisting of 20 men and 7 women among whom there were 15 persons who wore glasses. For all the 27 subjects, we collected 40 images for every subject in each illumination direction, so we got 4320 images in total. All images were color with 352*288 pixels in size (but in the experiment of this paper, all color images were transformed into gray level images). We call this face base IAE (Illumination and Expression) face base because the pose of each subject was almost unchangeable but the illumination direction and facial expression changed during the collecting process. Sample images in IAE face base are shown in Figure 4 (if the colors of the images were visible, you would tell that the illumination directions were obviously different). In Figure 4, the first row corresponds to the 0’ direction, the second row corresponds to the 30’ direction, the third row corresponds to the 60’ direction, and the fourth row corresponds to the 90’ direction, respectively. In the process of image collection, the head of a subject was not fixed and may sway slightly. Therefore the size of the face region and the illumination intensity may change slightly, which increased the difficulty of recognition. Experimental result, however, shows that good performance is achieved with the ensemble architecture although the imaging conditions were not strict.

### 4.2. Data Preprocessing

It is worth mentioning that the face images were normalized before they were presented to the system

**Figure 4: Samples in IAE Face Base**

for training or testing. The normalized version of a face image satisfies such constraints that the face could be appropriately cropped. The constraints include that the size of the image was fixed, the line between the two eyes was parallel to the horizontal axis, and the distance between the two eyes was set to a fixed value. The normalized face images in this paper satisfy the constraints in Figure 5.

A complete automatic face recognition system should include face detection and face recognition components. Because our emphasis was placed upon face recognition with changeable illumination conditions, we conducted face detection by manually

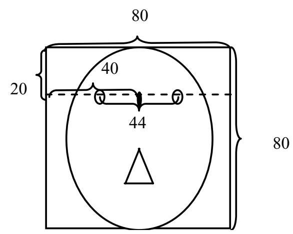


**Figure 5: Standard Face Image**

locating the eyes and then the row images were rotated and resized using the bilinear method. After that, face images were cropped to satisfy the constraints in Figure 5. In the end, the appropriately cropped face images were histogram equalized. Hence all the training and testing sets in this paper were normalized and histogram equalized gray level face images.

## 4.3. Data Sets and Training Method

We wanted to recognize 10 subjects in the IAE face base, and used another 10 subjects as a "*rejection*" subject, i.e., negative examples. We built 4 eigenspaces, each corresponding to an illumination direction. The training set for each eigenspace consisted of 270 face images among which each of the 27 persons in IAE face base contributed 10 images. Then we trained 4 BP neural networks with one hidden layer (the first layer), each corresponding to an illumination direction. The number of input units of each neural network was equal to the number of the dimensions of the corresponding eigenspace. Each neural network had 11 output units among which the first 10 units were corresponding to 10 subjects to be recognized and the last one was corresponding to the 10 subjects to be rejected, i.e., the negative examples. At the test phase, if a given face image was from the first 10 subjects, the system would tell the identity of the subject; if the image was from another 10 subjects to be rejected, the system would reject the image and mark it as "*unrecognizable*". The training set for each neural network was composed of 300 images with a corresponding illumination direction. Among the 300 images, each of the 10 persons to be recognized contributed 20 images, and each of the 10 persons to be rejected contributed 10 images. None of the 1200 (300*4) images have been used to build eigenspaces.

The training set for the architecture of neural network ensemble comprised of all the 1200 images used to train the first layer neural networks. The testing set consisted of 800 images, 200 images from each illumination direction. Among each 200 images, each of the 20 subjects contributed 10 images. None of the 800 images were used to build the eigenspaces and train the first layer neural networks.

Note that the parameters of neural networks have a great effect on the performance of neural networks and the number of units in the hidden layer is the most important one among all the parameters. So our experiment only took the number of units in the hidden layer into account. This number was adjusted several times and the best result was selected as the final evaluation of recognition performance.

## 4.4 Experimental Result

Taking different numbers (i.e. 20, 30 and 40) of dimensions for each eigenspace, we tested the ensemble architecture using the testing set with unknown illumination directions. For comparison, we also tested the conventional approach, which uses one neural network trained on a specific illumination direction. Specifically, we evaluated 4 conventional settings of the first layer neural networks, which were called 0' net, 30' net, 60' net and 90' net.

The experimental result is shown in Table 1. Each column stands for the testing images with a specific illumination direction, and each row shows the recognition ratio of each network or the ensemble architecture tested on the specific images.

Table 1 indicates that the ensemble architecture has achieved much higher performance (average 90.2%) without knowing or estimating the illumination direction. It is exciting that the ensemble architecture without illumination estimation even outperformed the best single neural network with known illumination direction (average 87.5%). The single neural network performed badly without known illumination direction (average 47.4%). It means the combination of the outputs from the neural networks of different illumination directions by a BP neural network can increase the performance.

The absolute recognition ratio is not very high and the recognition ratio on the 90' test set is 88.7%. It is mainly because, during the process of image collection,

**Table 1: Experimental Result**

| *Dimension Number* | *Neural Network* | *Testing Set* | | | | *Average* |
|---|---|---|---|---|---|---|
| | | *0' testing set* | *30' testing set* | *60' testing set* | *90' testing set* | |
| 20 | *0' net* | 165/200 | 125/200 | 67/200 | 25/200 | 382/800 |
| | *30' net* | 133/200 | 187/200 | 69/200 | 63/200 | 452/800 |
| | *60' net* | 60/200 | 87/200 | 171/200 | 71/200 | 389/800 |
| | *90' net* | 18/200 | 37/200 | 61/200 | 165/200 | 281/800 |
| | *Ensemble architecture* | 183/200 | 178/200 | 183/200 | 174/200 | 718/800 |
| 30 | *0' net* | 179/200 | 132/200 | 57/200 | 41/200 | 409/800 |
| | *30' net* | 138/200 | 187/200 | 79/200 | 92/200 | 496/800 |
| | *60' net* | 90/200 | 97/200 | 169/200 | 59/200 | 415/800 |
| | *90' net* | 10/200 | 10/200 | 33/200 | 180/200 | 233/800 |
| | *Ensemble architecture* | 183/200 | 182/200 | 175/200 | 180/200 | 720/800 |
| 40 | *0' net* | 161/200 | 125/200 | 47/200 | 43/200 | 376/800 |
| | *30' net* | 140/200 | 177/200 | 109/200 | 63/200 | 489/800 |
| | *60' net* | 63/200 | 100/200 | 182/200 | 46/200 | 391/800 |
| | *90' net* | 17/200 | 22/200 | 21/200 | 176/200 | 236/800 |
| | *Ensemble architecture* | 180/200 | 184/200 | 185/200 | 178/200 | 727/800 |

the heads of subjects were not fixed and they may sway slightly, so some of the images were blurry when collected by a camera.

## 5. Conclusion

We have described the illumination invariant face recognition method based on neural network ensemble. It can achieve higher recognition ratio without estimation of the illumination directions of the face images. The experiment, although only involving four illumination directions, has demonstrated the feasibility of our method. And it can be easily extended to a larger number of illumination directions.

In the real world, illumination is very complex indeed. The illumination directions and the illumination intensity are essentially arbitrary. Furthermore, a practical face recognition system should address both illumination and pose problems at the same time. It is interesting to generalize the work in this paper to recognize face images with arbitrary illumination and pose. A possible approach is to construct multiple eigenspaces from a large set of training images in a self-organizing way [12].

## Acknowledgements

The National Natural Science Foundation of P.R.China under grant No. 60273033 supported this research.

## References

[1] Zhao, W., Chellappa, R., Rosenfeld, A., and Phillips, P. J. Face recognition: a literature survey. http://citeseer.nj.nec.com / 374297.html, 2000.

[2] Brunelli, R., and Poggio, T. Face recognition: Features versus templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(10): 1042-1052, 1993.

[3] Laurenz, W., Jean-Marc, F., Norbert, K., and Christoph, v. d. M. Face Recognition by Elastic Bunch Graph Matching. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.19, 775-779, 1997.

[4] Edwards, G. J., Cootes, T. F., and Taylor, C. J. Face recognition using active appearance models. In: Proceedings of the 5th European Conference on Computer Vision, vol. 2, Freeburg, Germany, pp.581-595, 1998.

[5] Turk, M., and Pentland, A. Eigenfaces for recognition. Journal of cognitive neuroscience, 3(1): 71-86, 1991.

[6] Phillips, P. J., Wechsler, H., Huang, J., and Rauss, P. J. The FERET database and evaluation procedure for face-recognition algorithms. Image and Vision Computing, 16(5): 295-306, 1998.

[7] Pentland, A., Moghaddam, B., and Starner, T. View-based and modular eigenspaces for face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, 21-23, 1994.

[8] Huang, F. J., Zhou, Z-H., Zhang, H-J., and Chen, T. Pose invariant face recognition. In Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, pages 245-250, Grenoble, France, 2000. IEEE Computer Society.

[9] Hansen, L. K., and Salamon, P. Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(10): 993-1001, 1990.

[10] Opitz, D., and Shavlik, J. Actively searching for an effective neural network ensemble. Connection Science, 8(3-4): 337-353, 1996.

[11] Perrone, M. P., and Cooper, L. N. When networks disagree: Ensemble method for neural networks. In: Mammone, R. J. ed. Artificial Neural Networks for Speech and Vision. New York: Chapman & Hall, 126-142, 1993.

[12] Leonardis, A., Bischof, H., and Maver, J. Multiple eigenspaces. Pattern Recognition, (35): 2613-2627, 2002.