SCALE-ROBUST FEATURE EXTRACTION FOR FACE RECOGNITION

Zhifei Wang and Zhenjiang Miao

Institute of Information Science, Beijing Jiaotong University
No.3 Shangyuan Residence, Haidian District, Beijing 100044, P.R.China
phone:(86)010-51683149, email: {06112059, zjmiao}@bjtu.edu.cn

ABSTRACT

In video surveillance, the sizes of face images are very small. However, few works have been done to investigate scale-robust face recognition. Our experiments on appearance-based methods in different resolutions show that such methods as Neighboring Preserving Embedding (NPE) and Locality Preserving Projections (LPP) preserving local structure of data are less effective than the methods retaining global structure, for example, Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) under low-resolution condition. Based on these underlying phenomena, we propose a new graph embedding method named FisherNPE holding both global and local structures of data for scale-robust feature extraction. Experimental results on ORL and Yale database indicate that our method obtains good results on both low- and high-resolution images.

1. INTRODUCTION

Face recognition is a very active research field and previous works were mainly dealt with different complex conditions, for example, pose, illumination, and expression variations. However, image resolution remains a major concern, which is an important factor determining the performance of face recognition system [1]. In video surveillance applications, the sizes of interested faces are often small because of the distances between cameras and objects, which are often much smaller than the resolutions typically used in face recognition.

Currently, appearance-based method is one class of the most successful and well-studied techniques on face recognition. Although existing researches [2-7] have demonstrated that they were efficient in the high-resolution images, it keeps a puzzle that what the minimum resolution is to be detectable and recognizable in the face recognition systems. Some researches have shown that they could work well on low-resolution images. In [8], Kernel Correlation Feature Analysis (KCFA) method, which was proved to be highly successfully with low resolution images, outperformed PCA with a resolution of 8×8 pixels. PCA and LDA based classifiers [9] were used for recognition on face images of 16×12 pixels. In [10], they proposed a hypothesis that the requirement of resolution suitable for face registration was higher than for face recognition, and claimed that their system would work best with a resolution of 32×32.

Other works investigated the enhancement of low resolution images using face hallucination [11], which was used for obtaining high-resolution face images to get efficient facial features for recognition. It was an important research direction for solving low-resolution recognition problem. As a successful scale-invariant feature descriptor for object recognition, SIFT feature has been explored for face authentication in [12]. However, further studies are necessary to make sure whether SIFT features are suitable for describing face images.

In this work, we decide to analyze the effects image resolution bringing on the performance of face recognition. We present a scale-robust feature extraction method FisherNPE that combines NPE [5] local feature with LDA [3] global descriptor in the graph embedding framework. The NPE local feature aims at preserving local manifold structure of data and obtains good results on high-resolution images. The LDA global descriptor can be kept a good performance on low-resolution ones. Sufficient experiments are performed on ORL and Yale database with different resolutions, which show the effectiveness of the proposed method.

2. EFFECTS OF FACE IMAGE SCALE

We will first analyze the effects of image resolution on the performance of appearance-based methods for face recognition. Four representative appearance-based methods, PCA (Eigenfaces) [2], LDA (Fisherfaces) [3], LPP (Laplacianfaces) [4], and NPE (NPEfaces) [5], are used for research. As we all know, appearance-based methods perform badly under large variations in pose. Therefore, with the main purpose of investigating the effects of the resolution, ORL and Yale face databases are chosen for our experiments, since most of their face samples are frontal with small variations in pose. Six different resolutions are down-sampled from the original resolution using bi-cubic interpolation, and some results are shown in Figure 1.

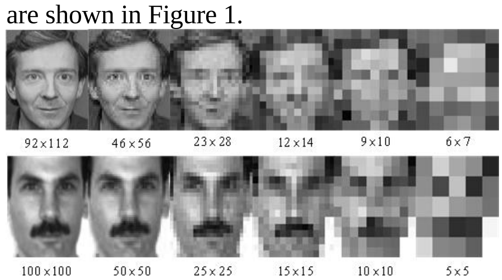

Figure 1 – Different resolution face images are used for experiments. (The upper row is from ORL, and the lower is from Yale database).

We design a group of experiments, which are performed with the same training data with original size and different low resolutions for testing. For each individual, m (=3, 4, 5) images are randomly selected for training and the rest for testing. In general, recognition rate varies with dimensions of feature vectors and parameters of systems. For given m (=5), we show the best results with different resolutions obtained by PCA, LDA, LPP and NPE with Nearest Neighbor classifier in Figure 2.

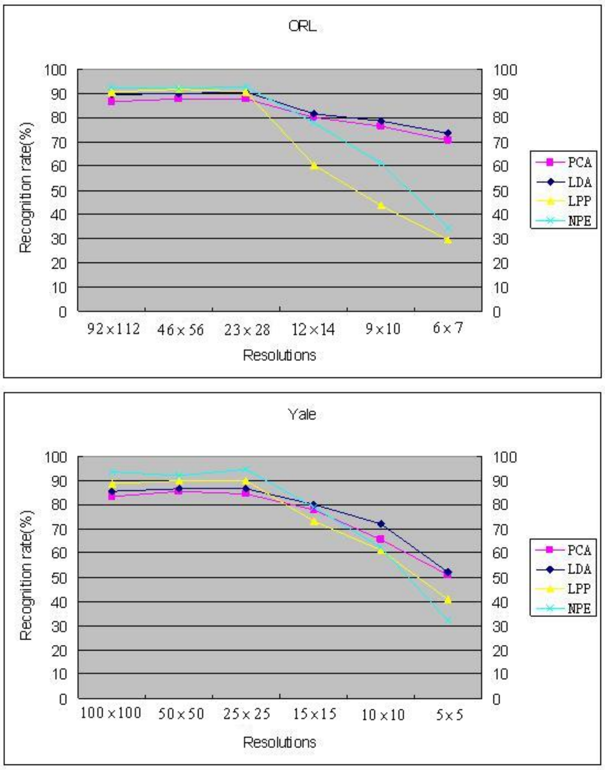

Figure 2 – Recognition accuracy versus image resolutions on ORL (top) and Yale (bottom) database with PCA, LDA, LPP, and NPE methods, respectively. X-axis represents for the six resolutions, respectively.

From the curve changing trend shown in Figure 2, we can obtain the following three important conclusions. First, we can observe that the threshold resolution is about 12×14 or 15×15, over the threshold resolution recognition rate remains stable roughly with resolution increasing, while under it recognition rate decreases seriously. The results are similar to the supposition proposed in [9], they proved that the discriminative information for recognition would increase with the resolution increasing and remain changeless when reaching at one certain resolution. Second, with the decreasing resolution LPP and NPE preserving local structure are affected more severely than PCA and LDA retaining global structure. Especially below threshold resolution, the resolution smaller, the effect greater. In [4-5], they pointed out that PCA and LDA aimed to discover global structure of Euclidean space and low-frequency information, while LPP and NPE aimed to discover local structure of face manifold and high-frequency information. However, low-resolution means making low-pass filter on the face image and destroying local structure of data, namely, high-frequency information. Finally, over threshold resolution, the methods preserving local structure obtain much better performance than global ones. Many works [4-7] have shown that NPE gains higher recognition accuracy than PCA and LDA on high-resolution images in most cases, however, the performance of LPP is no better than LDA's.

Based on the above-mentioned results and analyses, we decide to synthesize the advantages of global and local methods for scale-robust feature extraction. Linear framework of Graph Embedding (LGE) is proposed as

$$XWX^T a = \lambda XDX^T a \qquad (1)$$

The optimal a's are the eigenvectors corresponding to the maximum eigenvalue of eigen-problem of Eq.(1). The W and D are defined to characterize certain statistical or geometric properties of the data set. Detailed explanations were shown in [6-7]. Different choices of W and D will lead to some popular linear dimensionality reduction algorithms, which include LDA and NPE. They are the Linear extension of Graph Embedding problems $LGE(W^{LDA}, D^{LDA})$ and $LGE(W^{NPE}, D^{NPE})$, respectively. Therefore, we develop a new LGE method with the combination of LDA and NPE for achieving scale-invariant face recognition.

3. SCALE-ROBUST FEATURE DESCRIPTOR

Although LDA and NPE have the essential differences, they share some similar properties, for example, they are both linear and supervised dimensionality reduction algorithms. However, NPE can be also performed in unsupervised mode. For these reasons, we build a two-component vector fusing LDA descriptor preserving global structure and NPE descriptor retaining local structure. Thus, our vector is denoted as

$$F = \begin{bmatrix} (1-\omega)L \\ \omega N \end{bmatrix} \qquad (2)$$

Where L is the global LDA descriptor, N is the local NPE vector, and ω is the relative weighting factor.

Suppose we have m face image with the size of $h\times w$. Let $X=[x_1, x_2, ..., x_m]$ denote their vector representations with $\{x_i\}_{i=1}^m \subset \Re^n (n = h\times w)$.The graph embedding approach provides the mappings for the graph vertices in the training set and obtain the feature vectors. If we choose a linear function, $y_i = f(x_i) = a^T x_i$, i.e. we have $Y = X^T a$, the optimal feature vectors of LGE can be obtained by the following formula:

$$a = \arg\max \frac{Y^T WY}{Y^T DY} = \arg\max \frac{a^T XWX^T a}{a^T XDX^T a} \qquad (3)$$

Where W_{ij} denote the weight of the edge joining vertices x_i and x_j and $D_{ii} = \sum_j W_{ji}$. D is a diagonal matrix.

LDA [3] searches for the directions that are efficient for discrimination by maximizing the ratio between the interclass and intraclass scatters. The objective function is as follows:

$$L = \arg\max_{L^T S_W L = d} L^T S_B L = \arg\max_{L} \frac{L^T S_B L}{L^T S_W L} \quad (4)$$

$$S_B = \sum_{c=1}^{N} n_c (m^{(c)} - m)(m^{(c)} - m)^T \quad (5)$$

$$S_W = \sum_{c=1}^{N} \left(\sum_{i=1}^{n_c} (x_i^{(c)} - m^{(c)})(x_i^{(c)} - m^{(c)})^T \right) \quad (6)$$

Where n_c is the number of samples in the c th class, $m^{(c)}$ is the average vector of the c th class, m is the total sample mean vector, and $x_i^{(c)}$ is the i th sample in the c th class. In the Linear framework of Graph Embedding, [7] define W^{LDA} as:

$$W_{ij}^{LDA} = \begin{cases} \frac{1}{n_c}, \ if\ x_i \ and\ x_j \ are\ both\ belong \\ \quad to\ c-th\ class \\ 0, \quad otherwise \end{cases} \quad (7)$$

With the defined W^{LDA}, we can know that $D^{LDA} = I$. The global LDA vector L is corresponding to a in the LGE framework.

NPE [5] is a linear approximation to the LLE [13] algorithm. Let E denotes the weight matrix with E_{ij} ($E_{ij} = 0, if\ x_j \notin N_k(x_i)$, $N_k(x_i)$ denotes the set of k nearest neighbors of x_i) the weight of the edge joining vertices x_i and x_j, which can be computed by minimizing the following objective function,

$$\min_{\sum_j E_{ij}=1} \sum_i \left\| x_i - \sum_j E_{ij} x_j \right\|^2 \quad (8)$$

That is to say, solving the following generalized eigenvector problem,

$$XMX^T N = \lambda XX^T N \quad (9)$$

Where $M = (I-E)^T (I-E)$
$= I - (E + E^T - E^T E)$

Then we can define $W^{NPE} = E + E^T - E^T E$, so the minimization problem is changed into obtaining the vectors corresponding to the largest eigenvalues. It is easy to check that $D^{NPE} = I$. The local NPE descriptor N is corresponding to a in the LGE framework.

According to the above analyses, we can know some common properties between LDA and NPE. In NPE, W^{NPE} denotes the weights between local data points in a specified neighborhood. Similarly, W^{LDA} in LDA denotes the averaged weights between intraclass data points. Moreover, we all know that $D^{NPE} = D^{LDA} = I$. Therefore, LDA is a particular version of NPE to some extent. There exists large correlations between them. So, we integrate LDA with NPE embedding our scale-robust feature vector (2) into the LGE framework, and define the weight matrix W^F and D^F, then

$$XW^F X^T F = \lambda XDX^T F$$
$$\begin{cases} W^F = (1-\omega) W^{LDA} + \omega W^{NPE} \\ D^F = I \end{cases} \quad (10)$$

Where $W^F \in \Re^{m \times m}$ and $D^F \in \Re^{m \times m}$. FisherNPE descriptor F is corresponding to a in the LGE framework, which preserves both global and local structure of data to achieve scale-robust feature extraction.

Next, we will discuss about how to set the weighting factor ω, which is a function of resolution r (r denotes the size of input face image). That is to say, ω is different from different resolutions and then we can achieve scale-robust feature extraction. From the three conclusions about the effects of low resolution on the performance we obtain in section 2, there has a threshold resolution r_t. Below r_t, the performance degrades obviously. And up r_t, the performance keeps a constant high level. So we can empirically define ω as

$$\omega = f(r) = \begin{cases} r/r_o & if\ r \le r_t \\ 0.9 & otherwise \end{cases} \quad (11)$$

Where r_o is the original resolution of sample images. Note that r_t is a little different due to different databases. The weighting factor $\omega \in [0, 0.9]$ is an essential parameter in our FisherNPE method which controls the smoothness of feature descriptor. When $\omega \to 0$, FisherNPE will reduce to the LDA(Fisherfaces) methods. When $\omega \to 0.9$, FisherNPE will incline to the NPE(NPEfaces) methods.

4. EXPERIMENTS

In section 2, we have performed a few experiments on the effects of face image scale and obtained three important conclusions. In this section, we investigate mainly the use of FisherNPE on face analysis and recognition with varying resolutions, and compare it with LDA (Fisherfaces) and NPE (NPEfaces). All the preprocessing steps can be reviewed in section 2. Different resolution face images from ORL and Yale database used for training and testing are showed in Figure 1.

The first experiment for recognition is tested on the ORL (Olivetti Research Laboratory) face database [14]. It consists of a total of 400 face images with a total of 40 persons (10 samples per person). The images are all frontal and slight tilt of the head. We will perform three methods LDA, NPE, and FisherNPE on six different resolutions such as 92×112, 46×56, 23×28, 12×14, 9×10, and 6×7. For each individual, m (=3, 4, 5) images with original resolution are randomly selected for training and the rest with different resolutions with bi-cubic interpolation versions are used for testing. For each given m, the results are averaged over 5 random splits. In general, recognition rates vary with dimensions of the subspace. We select the optimal results from varying dimensions. Figure 3 shows the plot of recognition rates versus different resolutions in LDA, NPE and FisherNPE. The best results on the resolutions of 6×7 and 9×10 obtained in the optimal subspace are shown in Table1.

As can be seen, our FisherNPE method gains good results not only on the high-resolution images but also the low-resolution ones. Moreover, FisherNPE outperforms the other two methods across all the resolutions. Note that the good performances are obtained with different numbers of training samples and very low resolutions.

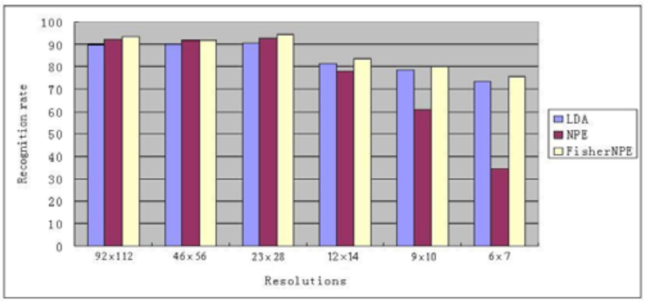

Figure 3 – Recognition rate vs image resolution on ORL database. (m=5)

Table 1 – Performance comparison on ORL with the resolutions of 6×7 and 9×10

6×7	3 Train	4 Train	5 Train
LDA	63.2%(79)	68.8%(119)	73.5%(159)
NPE	34.6%(30)	30.4%(30)	34.5%(30)
FisherNPE	64.3%(60)	70.8%(59)	75.6%(70)
9×10	3 Train	4 Train	5 Train
LDA	69.6%(79)	72.9%(119)	78.5%(159)
NPE	57.1%(30)	56.7%(30)	61%(30)
FisherNPE	70.4%(60)	73.8%(59)	80%(70)

Similar experiments are applied to the Yale database [15]. It contains 165 gray-scale images of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, with/without glasses, happy; left-light, w/no glasses, normal; right-light, sad, sleepy, surprised, and wink. We also perform the three methods on six different resolutions such as 100×100, 50×50, 25×25, 15×15, 10×10, and 5×5. A random subset with m (=3, 4, 5) images per individual are taken for the training set. The rest of the database is considered to be the testing set. For each given m, the results are averaged over 5 random splits. The experimental process is the same as before. Figure 4 and Table 2 show the recognition results respectively. The same good results as above are obtained. However, the performance of FisherNPE on the Yale is a little poorer than on the ORL, maybe due to the existence of illumination variation on the Yale.

As we all know, the performances of dimension reduction method vary with dimensions of the subspace. From our experiment results, when using different numbers of samples for training, the dimensions with good results are different. However, the dimensions with fine performance of the same method in different resolutions keep much changeless.

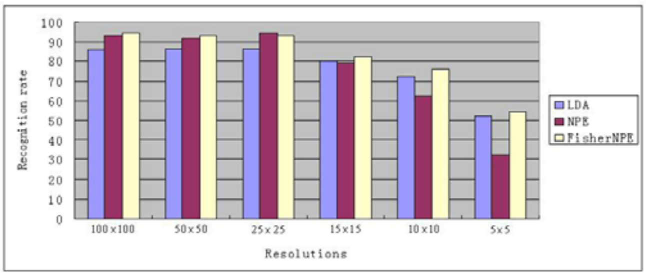

Figure 4 – Recognition rate vs image resolution on Yale database. (m=5)

Table 2 – Performance comparison on Yale with the resolutions of 5×5 and 10×10

5×5	3 Train	4 Train	5 Train
LDA	48.3%(29)	52.4%(44)	52.2%(59)
NPE	24.2%(30)	30.5%(25)	32.2%(30)
FisherNPE	51.4%(30)	52.3%(40)	54.4%(50)
10×10	3 Train	4 Train	5 Train
LDA	65%(29)	69.5%(44)	72.2%(59)
NPE	52.4%(30)	56.2%(25)	62.2%(30)
FisherNPE	66.7%(30)	73.3%(40)	75.6%(50)

5. CONCLUSIONS AND FUTURE WORK

In order to try solving low-resolution face recognition, we propose a scale-robust feature extraction method in Linear Graph Embedding framework. We call it as FisherNPE method, which combines LDA with NPE, preserves both local and global structures of data. Experiments on face recognition show that our method can obtain good performances not only on the high-resolution but also low-resolution images. The phenomenon of resolution threshold means that training sets with middle-resolution can be used for face recognition for better performance. However, compared with existing appearance-based methods, the improvement on performance of our method is limit. We have to develop more properties of preserving the structures of low-resolution face images. In the future, we will focus on testing our method on a large amount of face datasets for practical application. Furthermore, we will tackle images with different poses and illumination cases under low-resolution conditions. Other classifiers such as SVM could be considered for improving the performance of FisherNPE. Maybe kernelization and tensorization of our method will be explored for further researches.

6. ACKNOWLEGEMENTS

This work is supported by National 973 Key Research Program (2006CB303105), Innovative Team Program (IRT0707) of Ministry of Education of P.R.C, and Ph.D. Technology Innovation Fund of Beijing Jiaotong University.

7. REFERENCES

[1]W.Zhao, R.Chellappa, P.J.Phillips and A.Rosenfeld. "Face Recognition: A Literature Survey," *ACM Computing Surveys*, Vol.35, No.4, pp.399-458, Dec.2003.

[2]M.Turk and A.Pentland. "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*. Vol.3, No.1, pp.71-86, 1991.

[3]P.N.Belhumeur, J.P.Hespanha and D.J.Kriegman. "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. PAMI*, Vol.19, No.7, pp.711-720, Jul.1997.

[4]X.He and P.Niyogi. "Locality preserving projections," *In Proc. NIPS'03*, Vancouver, British Columbia, Canada, 2003.

[5]Xiaofei He, Deng Cai, Shuicheng Yan, and Hongjiang Zhang. "Neighborhood Preserving Embedding," *In Proc. ICCV'05*, Vol.2, Oct.2005, pp.1208-1213.

[6]Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, and Stephen Lin. "Graph Embedding and Extensions: A General Framework for Dimensionality Reduction," *IEEE Trans. PAMI*, Vol.29, No.1, pp.40-51, Jul.2007.

[7]Deng Cai, Xiaofei He, Yuxiao Hu, Jiawei Han, and Thomas Huang. "Learning a Spatially Smooth Subspace for Face Recognition," *In Proc. CVPR'07*, Jun.2007, pp.1-7.

[8]Ramzi Abiantun, Marios Savvides and B.V.K.Vijaya Kumar. "How Low Can You Go? Low Resolution Face Recognition Study Using Kernel Correlation Feature Analysis on the FRGCv2 dataset," *Biometric Consortium Conference, 2006 Biometric Symposium*, Aug.2006, pp.1-6.

[9]J.Wang, C.Zhang, and H.Shum. "Face image resolution versus face recognition performance based on two global methods," *In Proc. ACCV'04*, 2004.

[10]B.J.Boom, G.M.Beumer, L.J.Spreeuwers and R.N.J.Veldhuis. "The Effect of Image Resolution on the Performance of a Face Recognition System," *Control, Automation, Robotics and Vision, ICARCV'06*, Dec.2006, pp.1-6.

[11]S.Baker and T.Kanade. "Hallucinating faces," *In Proc. FGR'00*, Mar.2000. pp.83-89.

[12]M.Bicego, A.Lagorio, E.Grosso and M.Tistarelli. "On the use of SIFT features for face authentication," *In Proc. CVPRW'06*, Jun.2006, pp.35-41.

[13]Sam Roweis, and Lawrence K.Saul. "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, Vol.290, No.5500, pp.2323-2326, Dec.2000.

[14]AT&T Laboratories Cambridge. http://www.cl.cam.ac.uk/research/attarchive/facedatabase.html, 1994.

[15]Yale University. http://cvc.yale.edu/projects/yalefaces/yalefaces.html, 2002.