Face Recognition Based on Nonlinear Feature Approach

[1]Eimad E.A. Abusham, [1]Andrew T.B. Jin, [1]Wong E. Kiong and [2]G. Debashis
[1]Faculty of Information Science and Technology,
[2]Faculty of Engineering and Technology, Multimedia University (Melaka Campus),
Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia

Abstract: Feature extraction techniques are widely used to reduce the complexity high dimensional data. Nonlinear feature extraction via Locally Linear Embedding (LLE) has attracted much attention due to their high performance. In this paper, we proposed a novel approach for face recognition to address the challenging task of recognition using integration of nonlinear dimensional reduction Locally Linear Embedding integrated with Local Fisher Discriminant Analysis (LFDA) to improve the discriminating power of the extracted features by maximize between-class while within-class local structure is preserved. Extensive experimentation performed on the CMU-PIE database indicates that the proposed methodology outperforms Benchmark methods such as Principal Component Analysis (PCA), Fisher Discrimination Analysis (FDA). The results showed that 95% of recognition rate could be obtained using our proposed method.

Key words: Feature extraction, LLE, FDA, LFDA, manifold learning

INTRODUCTION

During the past decades, Face Recognition (FR) has become an active area of research in computer vision, neuroscience and psychology. Two issues are central to face recognition algorithms (i) feature selection for face representation and (ii) classification of a new face image based on the chosen feature representation. In face recognition, such as Eigenfaces by Matthew and Pentland[1] and Eigenfaces proposed by Belhumeur and Hespanha [2] are two well-known linear projection methods for data reduction and feature extraction under the unsupervised and supervised learning settings, respectively. Many state-of-the-art FR methods are built on these two techniques or their variants. Eigenface technique, one of the most successful face recognition methods based on linearly projecting the image space to a low dimensional subspace it finds the optimal projection directions that maximally preserve the data variance. Eigenface technique has a limitation it does not give the most discriminating features for separating different face classes. Fisherface method is based on Fisher's Linear Discriminant and produces well separated classes in a low dimensional subspace, even under severe variation in lighting and facial expressions. Extensive experimental results conducted by Belhumeur and Hespanha[2] demonstrate that the proposed Fisherface method has error rates that are lower than those of the Eigenface technique for tests on the Harvard and Yale Face Databases.

Fisher Discriminant Analysis (FDA) based algorithms suffer from the so-called small sample size problem(SSS)[3] which exists in high dimensional pattern recognition tasks where the number of available samples is smaller than the dimensionality of the samples. Due to this problem, many variants of the original FDA algorithm have been proposed for face recognition. Belhumeur and Hespanha[2] has proposed to perform PCA first before applying LDA in the PCA-based subspace also known as PCA+LDA; but he found that Fisherface does not gave better performance than Eigenface in many cases, The reason is generalization ability of Fisherface can be degraded since some principal components with small eigenvalues correspond to high-frequency components can cause latent noise.

Recently, Numerous methods have been proposed to overcome this difficulty, Direct LDA (D-LDA) method has been presented in[3,4]. At the core of the direct LDA algorithm lies the idea of simultaneous diagonalization, As the name suggests, it tries to find a matrix that simultaneously diagonalizes both between class and within class scatter matrixes.LDA/GSVD method which is a generalization of LDA based on the Generalized Singular Value Decomposition (GSVD)[5].

Corresponding Author: Eimad Eldin Abdu Ali Abusham, Faculty of Information Science and Technology, Multimedia University (Melaka Campus),Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia Tel:+60126348120 Fax:+60623118840

LDA/GSVD extends the applicability to cases that classical discriminant analysis cannot handle. LDA/QR method combining GSVD and QR decomposition proposed by Ye and Li[6] which aims to overcome the singularity problems of classical LDA, while achieving efficiency and scalability simultaneously. These methods proposed to avoid the shortcomings existing in traditional solutions to the SSS problem. While LDA-based methods have been demonstrated to perform well on many pattern classification applications, their performance is unsatisfactory when applied to problems described complex nonlinear variation of face images. It is reasonable to assume that a better solution to nonlinear problem could be achieved using nonlinear methods,

Zhang, Li and Wang presented a new algorithm, based on Manifold Learning and Nonlinear subspace. They have divided manifold learning into four major classes: projection methods, generative methods, embedding methods and mutual information methods[7,8]. The core of the first class is to find some surfaces passing through the middle of data, such as the principal curves[9,10]. For the second class adopts generative topology models[11,12] and hypothesizes that observed data are generated from the evenly spaced low-dimensional latent nodes. Then the mapping relationship between the observation space and the latent space can be employed for classification and visualization. Resulting from the inherent insufficiency of the adopted Expectation and Maximization algorithms, nevertheless, the generative models fall into local minimum easily and also have slow convergence rates. For the third class is generally divided into global and local embedding algorithms. ISOMAP[13], as a global algorithm, presumes that isometric properties (e.g. geodesic) should be preserved in both the observation space and the intrinsic embedding space in the affine sense. For Local preservation they have presented Locally Linear Embedding (LLE)[14] and Laplacian Eigenamp[15] focus on the preservation of local neighbor structure. For the fourth category, it is assumed that the mutual information is a measurement of the differences of probability distribution between the observed space and the embedded space, as in stochastic nearest neighborhood[16] and manifold charting[17].

For face recognition, classical dimensionality reduction methods include Eigenface, Fisher Discriminant Analysis and Local Feature Analysis (LFA), etc. The linear methods have their limitations. Firstly, they cannot reveal the intrinsic distribution of a given data set. Secondly, if there are changes in pose, facial expression and illumination, the projections may not be appropriate and the corresponding reconstruction error may be much higher. To overcome these

problems, we propose a new method combing the advantages of linear and nonlinear methods, which is the combination of locally linear embedding and local fisher discriminant analysis. Locally linear embedding, which is able to do nonlinear dimensionality reduction in an unsupervised way. A disadvantage of LLE algorithm is that mapping of test samples is difficult for computation cost of eigenmatrix. Thus our novel approach manages to overcome these problems.

BRIEF REVIEW OF LLE AND LFDA

Locally linear embedding: Saul and Roweis[14] proposed a conceptually simple yet powerful method for nonlinear mapping namely, Locally Linear Embedding (LLE). LLE establish the mapping relation between the observed data and the corresponding low-dimensional one, the locally linear embedding algorithm is used to obtain the low-dimensional data $Y(Y \in \mathbb{R}^d)$ of the training set $X(X \in \mathbb{R}^N, N \gg d)$. Then the data set (X,Y) is used for modeling the subsequently mapping relationship.

The main principle of LLE algorithm is to preserve local neighborhood relation in both the embedding space and the intrinsic one. Each sample in the observation space is a linearly weighted average of its neighbors. The basic LLE algorithm is described as follows:

Step 1:

$$\varepsilon(W) = \sum |x_i - \sum W_{ij} x_j|^2 \qquad (1)$$

Which adds up the squared distances between all the data points and their reconstructions. x_{ij} are the neighbors of x_i enforcing $W_{ij} = 0$. if x_j does not belong to this.

Step 2: Assigning a weight to every pair of neighboring points. The weights representing contributions to the reconstruction of a given point from its nearest neighbors can be found by solving the optimization task.

$$\Phi(Y) = \sum |y_i - \sum W_{ij} y_j|^2 \qquad (2)$$

subject to constraints $W_{ij} = 0$, if x_i and x_j are not neighbors and $\sum_{j=1}^{N} W_{ij} = 1$.

Step 3: The algorithm is to approximate the nonlinear manifold around sample x_i linear hyberplane that passes through its neighbors $\{x_{i1}, \ldots, x_{ik}\}$. considering that the objective $\Phi(Y)$ is invariant to translation in Y, constrain

term $\sum_i y_i = 0$ is added to step 2. Moreover. The other term $\sum_i y_i y_i^T / n = I$ is to avoid the degenerate solution of Y = 0. Hence; the step 2 is transformed to the solution of eigenvectors decomposition which can be seen as follows:

$$\tilde{Y} = \arg\min \Phi(Y) = \| y_i - \sum_{j=1}^{K} W_{ij} y_{ij} \|^2$$

$$= \arg\min_Y \| (I - W)Y \|^2 = \arg\min_Y Y^T (I - W)(I - W)Y \qquad (3)$$

The optimal solution of $\tilde{Y}$ in formula (3) is the smallest eigenvector of matrix (I-W) (I-W). With respect to the constraint conditions, the eigenvalue, which is zero, need to be removed. So we need to compute the bottom (d + 1) eigenvectors of the matrix and discard the smallest eigenvectors considering constraint term.

Local Fisher Discriminant Analysis (LFDA): To overcome the problems of undesired behavior of FDA and its variants caused by the globality when evaluating the within-class scatter and the between-class, Sugiyama (2006) proposed a method allow attaining between-class separation and within-class local structure preservation at the same time.

Let $x_i \in \mathbb{R}^D$ (I = 1, 2, …, n) de D-dimensional samples and $y_i \in \{1, 2, …, C\}$ be associated class labels, where n in the number of samples and C is the number of classes. Let us define the local within-class scatter matrix $\overline{S}^{(w)}$ and the local between-class scatter matrix $\overline{S}^{(b)}$ as follows.

$$\overline{S}^{(w)} = \frac{1}{2} \sum_{i,j=1}^{n} \overline{W}_{i,j}^{(w)} (x_i - x_j)(x_i - x_j)^T, \qquad (3)$$

$$\overline{S}^{(b)} = \frac{1}{2} \sum_{i,j=1}^{n} \overline{W}_{i,j}^{(b)} (x_i - x_j)(x_i - x_j)^T, \qquad (4)$$

where,

$$\overline{W}_{i,j}^{(w)} = \begin{cases} A_{i,j}/n_c & \text{if } y_i = y_j = c, \\ 0 & \text{if } y_i \neq y_j, \end{cases}$$

$$\overline{W}_{i,j}^{(w)} = \begin{cases} A_{i,j}(1/n - 1/n_c) & \text{if } y_i = y_j = c, \\ 1/n & \text{if } y_i \neq y_j, \end{cases}$$

According to the affinity $A_{i,j}$ we weight the values for the sample pairs in the same class.

This means that far apart sample pairs in the same class have less influence on $\overline{S}^{(w)}$ and $\overline{S}^{(b)}$. LFDA do not weight the values for the sample pairs in different classes since we want to separate them from each other irrespective of the affinity in the original space.

Note: If the affinity value $A_{i,j}$ is set to 1 for all sample pairs, $\overline{S}^{(w)}$ and $\overline{S}^{(b)}$, respectively and LFDA is reduced to the original FDA. Therefore, LFDA may be regarded as a natural localized variant of FDA.

Using $\overline{S}^{(w)}$ and $\overline{S}^{(b)}$, we define the LFDA transformation matrix (T) T_{LFDA} as:

$$T_{LFDA} = \arg\max_{T \in \mathbb{R}^{d \times m}} \text{tr}\left((T^T \overline{S}^{(w)} T)^{-1} T^T \overline{S}^{(b)} T\right). \qquad (5)$$

and

$$\overline{S}^{(m)} = \overline{S}^{(w)} + \overline{S}^{(b)}. \qquad (6)$$

Properties of LFDA: The interpretation of LFDA in terms of the 'pointwise scatter'. $\overline{S}^{(w)}$ Can be expressed as:

$$\overline{S}^{(w)} = \frac{1}{2} \sum_{i=1}^{n} \frac{1}{n_{yi}} \overline{P}_i^{(w)},$$

Where n_{yi} is the number of samples in the class to which the sample x_i belongs and $\overline{P}_i^{(w)}$ is the pointwise local within-class scatter matrix around x_i:

$$\overline{P}_i^{(w)} \equiv \sum A_{i,j}(x_j - x_i)(x_j - x_i)^T.$$

Therefore, minimizing $\overline{S}^{(w)}$ corresponds to minimizing the weighted sum of the pointwise local within-class scatter matrices over all samples. $\overline{S}^{(b)}$ can also be expressed in a similar way as:

$$\overline{S}^{(b)} = \frac{1}{2} \sum_{i=1}^{n} \left(\frac{1}{2} - \frac{1}{n_{yi}} \right) p_i^{(w)} + \frac{1}{2n} \sum_{i=1}^{n} p_i^{(b)}, \qquad (7)$$

where $p_i^{(b)}$ is the pointwise between-class scatter matrix around x_i:

$$P_i^{(b)} \equiv \sum_{j:y_j \neq y_i} (x_j - x_i)(x_j - x_i)^T.$$

$P_i^{(b)}$ does not include the localization factor $A_{i,j}$. in Eq.

13 implies that maximizing $\overline{S}^{(b)}$ corresponds to minimizing the weighted sum of the pointwise local within-class scatter matrices and maximizing the sum of the pointwise between-class scater matrices.

In practice, Sugiyama[18] proposed determining the LFDA transformation matrix T_{LFDA} as follows. Rescale the generalized eigenvectors $\{\overline{\varphi}_k\}_{k=1}^d$ so that:

$$\overline{\varphi}_\kappa \overline{S}^{(w)} \overline{\varphi}_{k`} = \begin{cases} 1 & \text{if } k=k`, \\ 0 & \text{if } k \neq k`. \end{cases} \tag{8}$$

Note that this rescaling is often automatically carried out by an eigensolver. Then we weight each generalized eigenvector by the square root of its associated generalized eigenvalue, that is:

$$T_{LFDA} = (\sqrt{\overline{\lambda}_1 \overline{\varphi}_1} \,|\, \sqrt{\overline{\lambda}_2 \overline{\varphi}_2} \,|...|\, \sqrt{\overline{\lambda}_r \overline{\varphi}_r}), \tag{9}$$

Where $\overline{\lambda}_1 \geq \overline{\lambda}_2 \geq ... \geq \overline{\lambda}_d$. This weighting scheme weakens the influence of minor eigenvectors.

PROPOSED METHOD (LLELFDA)

Introduction: For PCA the extracted features are global features for all face classes and thus may not be optimal for discriminating one face class from the others. LDA and its variants although successful in many cases, linear methods fail to deliver good performance when face patterns are subject to large variations in viewpoints, which results in a highly non-convex and complex distribution. To cope with drawbacks, it is reasonable to use nonlinear methods when we dealing with subject that has large variations in viewpoints. In this paper we propose a new method to overcome the limitation of PCA and overcome the weakness of FDA by integrating Locally Nonlinear Embedding (LLE) proposed by Saul and Roweis[14] and Local Fisher Discriminating Analysis (LFDA) recently reported Sugiyama[18], LLE is Manifold learning approaches which is aimed to discover the intrinsical low dimensional variables from high dimensional nonlinear data then LFDA algorithm is used to project samples into discriminant space to attain between-class

separation and within class local structure preservation at the same time.

Llelfda: Here we want to perform LFDA in the LLE low dimensional embedding space $\mathbb{R}^d$. First, LLE to transform the input space $\mathbb{R}^N$ into an m-dimensional space. Pattern x in $\mathbb{R}^N$ is transformed to be LLE based feature vector y in $\mathbb{R}^d$. We can obtain the LLE transformed feature vector $Y = (y_1, y_2, ..., y_m)^T$ by (2)

Second, from transformed feature vector $Y = (y_1, y_2, ..., y_m)^T$ obtained by LLE we construct the local between-class and local within-class scatter matrices $\overline{S}^{(w)}$ and $\overline{S}^{(b)}$ to obtain the T_{LFDA}, as in (4,5)

$$T_{LFDA} = \arg \max_T \left[tr((T\overline{S}^{(w)}T)^{-1} T\overline{S}^{(b)}T) \right] \tag{10}$$

LFDA solution can be obtained just by solving a generalized eigenvalue problem.

$$\overline{S}^{(b)} \varphi = \lambda \overline{S}^{(w)} \varphi$$
$$T_{FLDA} = (\overline{\varphi}_1, \overline{\varphi}_2, ..., \overline{\varphi}_R) \tag{11}$$

RESULTS AND DISCUSSION

Results: In this section we summarize the results obtained by our proposed method and compare it with benchmark methods. The results were obtained as follows. The database images were divided into training and a testing set (several random partitions were tried in every experiment). In this study, CMU-PIE is tested. CMU-PIE (pose, illumination and expression) database from CMU-PIE[19]. In our experiment, the facial portion of each original image was cropped and resized pixels and preprocessed by histogram equalization (HE).

CMU-PIE database contains 68 subjects with 41,368 face images as a whole. 13 synchronized cameras and 21 flashes, under varying pose, illumination and expression, captured the face images. Fig. 1. shows some of the faces with pose, illumination and expression variations in the PIE database (normal and preprocessed (HE) images of the same subject. In our experiments the proposed method was tested on subset of CMU-PIE database. This subset includes 1360 images of 68 subjects/individuals for training and testing set without overlapping (each individual has twenty images 20×68). The size of each cropped image in all the experiments is 32×32 pixels, with 256 gray levels per pixel. Thus, a 1,024 dimensional vector in image space represents each image.

First, training samples are mapped into low dimensional embedding space and then LFDA algorithm is used to project samples into discriminant space for enlarging between class distances and within-

Fig. 1: Sample face images of one subject used in the experiments, original images (left), HE (right)

Table 1: Cumulative match characteristic vs. rank

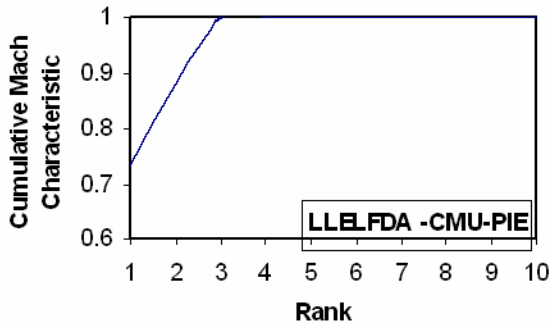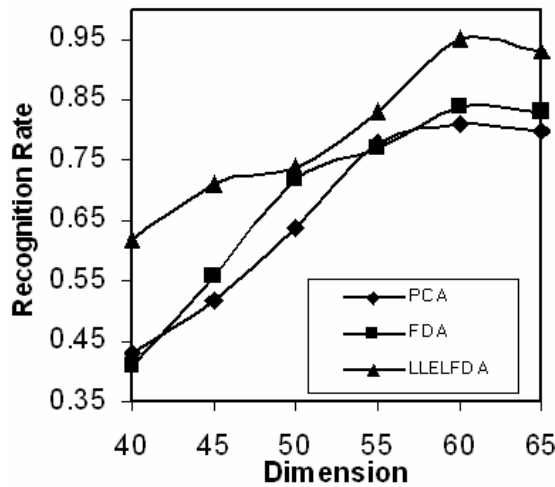|	Rank1	Rank2	Rank3	Rank4	Rank6	Rank7	Rank8	Rank9	Rank10
LLELFDA	0.735	0.882	1	1	1	1	1	1	1

Fig. 2: CMC for CMU-PIE database

Fig. 3: Recognition Rate vs. Dimension

class local structure is preserved. The performance of FLDA based on the number of principal components in LLE. To measure the performance we have used Cumulative Match Characteristic vs. Rank. CMC is a method of showing measured accuracy performance of our proposed method operating in the closed-set identification task. Images are compared and ranked based on their similarity. The CMC shows how often the individual's template appears in the ranks (1, 2, 3, 4,

etc.), based on the match rate. Fig. 2. shows the performance of LLELFDA on CMC Curve.

Table 2: The average recognition rates (%) of PCA, FDAand LLELFDA across ten tests and four dimensions (50,55,60,65)

PCA	FDA	LLEFLDA
75.75	79.00	86.25

The proposed LLELFDA is tested and compared with popular benchmark method PCA and FDA. Simple classifier nearest neighbor classifier (NN) is employed in the experiments. From Fig. 3. We can see that the performance of LLELFDA is consistently better than PCA and FDA. PCA and FDA can only utilize the linear information. Whereas, LLELFDA is able to make use of nonlinear discriminant information and then perform further linear discriminant on the transform space obtain by LLE, which turn out to be complimentary for achieving a better result.

The result indicates that FDA and its invariant are really helpful for improving the performance of face recognition. FDA still has the constraint of $T_{FDA}^T S^{(w)} T_{FDA} = I_r$, where I_γ is the identity matrix on R^γ and T is the transformation matrix of FDA for between-class scatter matrix. LLELFDA effectively combines the ideas of LLE and LFDA. LFDA maximizing between-class separability and preserving the within-class local structure at the same time without sharing the limitation of FDA and can be employed for dimensionality reduction into any dimensional spaces. This consider as a significant improvement over the original FDA.

For more consistency, we have tested the proposed method on different training set and repeat the experiments ten times. Each time the training set is selected randomly. Correspondingly, the testing sets are different as well (number of subject in the raining and testing set same as in CMU-PIE experiment)

For each experiment, four different dimension (50,55,60,65) are chosen, where we obtain good recognition rates; as we can see from Fig. 3. For PCA, FDA and LLELFDA, the average recognition rates across ten tests and four dimensions are listed in Table 2.

DISCUSSION

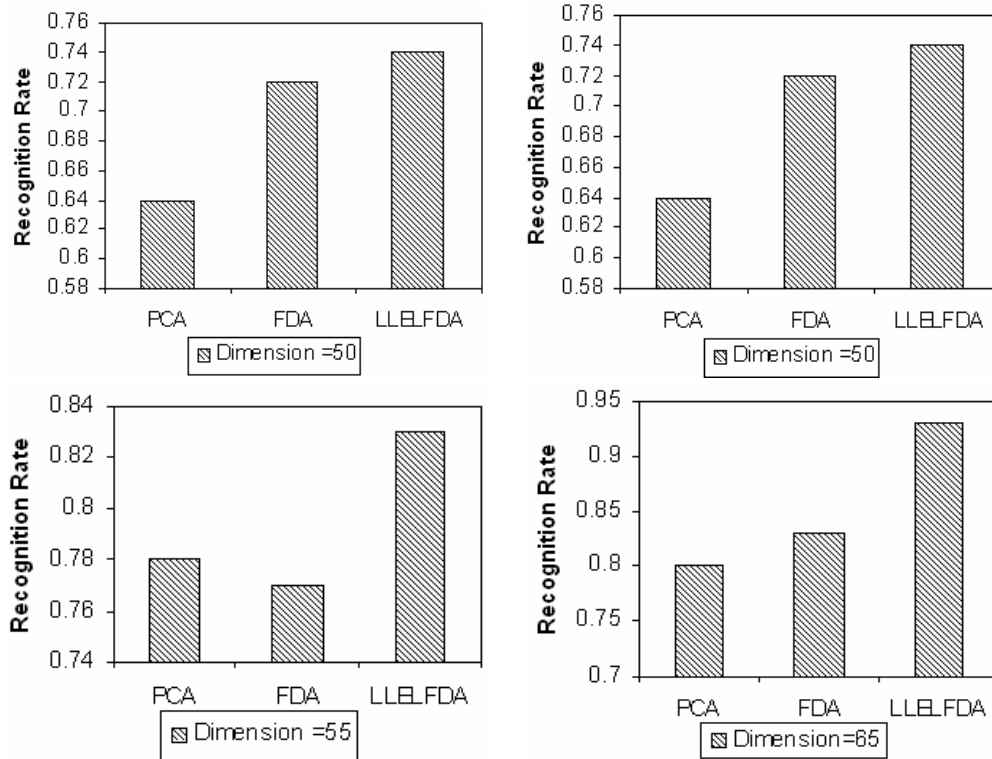Literature surveys and previous studies stated that if linear classifier doesn't work well, then there are at

Fig. 4: The averages recognition rates of PCA, FDA and LLELFDA across 10 tests (dimension 50,55,60,65)

least two potential reasons for this:1) regularization was not done well or no robust estimators were used. 2) Intrinsically nonlinear. Since our dataset is high dimensional data and the nature of face images are nonlinear, then it is recommended to use a linear classifier in the appropriate nonlinear feature space. Our proposed method is utilizing the advantages of nonlinear and linear method in one frame; we have used LLE for mapping high dimensional data to a low dimensional observed space. The most attractive virtues of LLE are that there are only two parameters to be set and the computation can avoid converging to a local minimum. Once the data are mapped to low dimensional space, LFDA run on the reduced samples. LFDA is used to highly separate embedding space obtained by LLE, it give more separate embedding than FDA and its variants. The reason is, FDA can be regarded as maximizing between class scatter under the constraint of keeping within class scatter to a certain level and this restriction is quite hard since we deal with multimodal face images, this will lead to face images of certain multimodal class to be merged into a single class. Thus FDA and its variants results in less separate

embedding. LFDA does not require multimodal of certain class to map into a single class. As a result, more degree of freedom is left for increasing separability and thus highly separate embedding can be obtained. In our experiments, we study the effect of the number of reduced dimensions ranges from 50 to 65 for our dataset. The recognition results on the unified and benchmark method is shown in Fig. 3., where the horizontal axis is the number of reduced dimensions and the vertical axis is the recognition accuracy. We can observe that the accuracy tends to increase when the number of reduced dimension increases until it near or reached the rank of between class scatter matrixes. We applied our method to ten datasets from CMU-PIE database and compared with benchmark method PCA and FDA in terms of recognition accuracy. Four different dimensions (50, 55, 60, 65) were chosen, where we obtain good recognition rates. The averages recognition rates of PCA, FDA and LLELFDA across 10 tests is shown in Fig. 4. The result summarized in Table 2. The main observations from table 2. Our method is competitive and outperforms the benchmark methods.

CONCLUSION

In this study, we have proposed a novel approach method called LLELFDA based on nonlinear feature extraction. The proposed method LLELFDA can be used to carry out discriminant analysis in double discriminant subspaces (nonlinear and linear), it can also outperform the benchmark methods PCA and FDA algorithms because, LLELFDA also improve the discriminating power of the extracted features by maximize between-class while within-class local structure is preserved. In addition to that, our method overcome the limitation of linear method where the constrain on between class is less compare to FDA and its variants. Extensive face recognition experiments based on the CMU-PIE face databases show that LLELFDA significantly outperforms PCA and is also superior to FDA.

REFERENCES

1. Matthew, A.T. and A.P. Pentland, 1991. Eigenfaces for Recognition. J. Cogn. Neurosci., 3: 71-86.
2. Belhumeur, P.N., J.P. Hespanha and D.J. Kriegman, 1997. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intelligence, 19: 711-720.
3. Chen, L.F., M.L. Hong-Yuan, K. Ming-Tat, L. Ja-Chen and Y. Gwo-Jong, 2000. A new LDA-based face recognition system which can solve the small sample size problem. Pattern Recognition, 33: 1713-1726.
4. Yu, H. and J. Yang, 2001. A direct lda algorithm for high-dimensional data with application to face recognition. Pattern Recognition, 34: 2067–2070.
5. Howland, P. and H. Park, 2004. Generalizing discriminant analysis using the generalized singular value decomposition. IEEE Trans. Pattern Anal. Mach. Intell., 26: 995-1006.
6. Ye, J. and Q. Li, 2005. A two-stage linear discriminant analysis via QR-decomposition. IEEE Trans. Pattern Anal. Mach. Intell., 27: 929-941.
7. Zhang, J., J. Wang, 2004. Manifold Learning. Neural Networks and Applications. Zhou, Z.H. and C.G. Cao (Eds.), Tsinghua University Press, pp: 172-207.
8. Zhang, J., S.Z. Li and J. Wang, 2004. Manifold Learning and Applications in Recognition. In: Intelligent Multimedia Processing with Soft Computing, Tan, Y.P., K.H. Yap and L. Wang (Eds.), Springer-Verlag, Heidelberg.
9. Hastie, T. and W. Stuetzle, 1988. Principal Curves. J. Am. Stat. Assoc., 84: 502-516.
10. Kégl, B., A. Krzyzak, T. Linder and K. Zeger, 2000. Learning and Design of Principal Curves. IEEE Trans. Pattern Anal. Mach. Intelligence., 22: 281-297.
11. Bishop, C.M., M. Sevensen and C.K.I. Williams, 1998. GTM: The generative Topographic Mapping. Neural Computat., 10: 215-234.
12. Chang, K. and J. Ghosh, 2001. A Unified Model for Probabilistic Principal Surfaces. IEEE Trans. Pattern Anal. Mach. Intelligence, 23: 22-24.
13. de Silva, T.J.B. and J.C. Langford, 2000. A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science, 290: 2319-2323.
14. Roweis, S.T. and K.S. Lawrance, 2000. Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science, 290: 2323-2326.
15. Belkin, M. and P. Niyogi, 2003. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Neur. Computat., 15: 1373-1396.
16. Hinton, G. and S. Roweis, 2002. Stochastic Neighbor Embedding. Neural Information Proceeding Systems: Natural and Synthetic, Vancouver, Canada.
17. Brand, M. Merl, 2002. Charting A Manifold. Neural Information Proceeding Systems: Natural and Synthetic, Vancouver, Canada.
18. Sugiyama, M., 2006. Local Fisher Discriminant Analysis for Supervised Dimensionality Reduction, in Proceedings of 23rd International Conference on Machine Learning, Pittsburgh, PA, pp. 905-912.
19. Sim, T., S. Baker and M. Bsat, 2001. The CMU pose, illumination and expression (PIE) database of human faces. Technical Report CMU-RI-TR-01-02, Carnegie Mellon University.