Proceedings of the World Congress on Engineering 2008 Vol I
WCE 2008, July 2 - 4, 2008, London, U.K.

Face Recognition using PCA and LDA with Singular Value Decomposition(SVD) using 2DLDA

Neeta Nain *, Prashant Gour, Nitish Agarwal, Rakesh P Talawar, Subhash Chandra

Abstract—**Linear Discriminant Analysis(LDA) is well-known scheme for feature extraction and dimension reduction. It has been used widely in many applications involving high-dimensional data, such as face recognition. In this paper we present a new variant on Linear Discriminant Analysis (LDA) for face recognition by reducing dimensions of input data using matrix representation and after that using singular value decomposition to reduce dimensions of scatter matrix. Experiments on ORL face database shows the effectiveness of our proposed algorithm and results compared with other LDA based methods shows that the proposed scheme gives comparatively better results than previous methods in terms of recognition rate and reduced time complexity.**

***Keywords:* Face recognition, Principal Component Analysis (PCA),Linear Discriminant Analysis (LDA),Relevance Weighted LDA, LDA/QR and Singular Value Decomposition (SVD)**

1 Introduction

Recent years have witnessed a growing interest of researchers in the field of face recognition Problem. This problem is studied in different fields with different point of views. There are several application areas where face recognition in our life such as identification of person using Credit cards, Passport check, Criminal investigations etc. Principal Component Analysis (PCA) [1] is one of the most popular appearance-based methods used mainly for dimensionality reduction in compression and recognition problem. PCA is known as Eigenspace Projection which is based on linearly Projection the image space to a low dimension feature space that is known as Eigenspace. It tries to find Eigen vectors of Covariance matrix that corresponds to the direction of Principal Components of original data.

Another powerful dimensionality reduction technique is Linear Discriminant Analysis (LDA) [2]. LDA is a classical method for feature extraction and dimensionality reduction which has been widely used in several classification problems. The objective of LDA is to find out the optimal transformation matrix so the ratio of between class scatter matrix and within class scatter matrix reaches to its maximum. However, in face recognition problems a critical issue using LDA is the small sample size (SSS) [3] problem. This problem arrives when there is are small number of training samples but the dimension of the feature space is large. This means that the within class scatter matrix would tend to be a singular matrix and so the algorithm fails. To overcome this problem some enhancements in classical LDA have been proposed. Among them most popular one is using PCA as a pre-processing step and then perform LDA so dimensionality reduction occur during PCA phase. But intermediate dimension reduction may lead to information lost. Another method called LDA/QR [4] also has been presented to overcome the SSS problem by performing QR decomposition. Due to QR decomposition dimensions of scatter matrices reduces because of orthonormal matrix produced by decomposition. But it has been observed that the class separability criteria proposed by classical LDA does not maximize the classification accuracy. It results in preserving the distance of previously well separated classes but an overlapping of neighbor classes also occurs. To overcome this problem an extended scheme is also proposed [2] by applying weighting function in the estimation of scatter matrices. M. Loog [2] has presented an LDA enhancements algorithm namely relevance weighted LDA (RW-LDA) by replacing the unweighted scatter matrices through the weighted scatter matrices in the classical LDA method.

This method alone is not applied because of singularity problem so we propose a new method LDA-RW with SVD (Singular value decomposition) using 2DLDA [5] as preprocessing step which overcomes the singularity problem. We are representing input data in the matrix form and by calculating left and right transformation matrices, reduce the input data dimensions and then we use orthonormal matrix created by SVD so that the dimensions of scatter matrices reduces and the singularity problem does not arise. The paper is organized as: Section 2 details the existing techniques. Section 3 describes the proposed

*Date of Submission: 1^{st} of April, 2008. Address of the corresponding author: Department of Computer Engineering, Malaviya National Institute of Technology, Jaipur–302017, India. Telephone/Fax No.: +91 141 2529140/+91 141 2713285

method. The effectiveness of our method is compared by performing tests using well known ORL database. Tests are carried out and results are compared with other LDA based methods, such as popular Fisherface method and the more recent LDA/QR method in Section 2.4. Finally, Section 5 concludes the proposed algorithm.

2 Review of Previous Approaches

This section details the PCA and the previous LDA literatures. PCA emphasizes on the data, whereas, LDA emphasizes on finding the relationships between different classes.

2.1 Principal Component Analysis(PCA)

PCA is a classical feature extraction widely used in the area of face recognition to reduce the dimensionality. PCA seeks to find the vectors that best describe the data in terms of reproducibility; however these vectors may not include the effective information for classification, and may eliminate discriminative information.

PCA aims to find the eigenvalues of the covariance matrix C,

$$c = \frac{1}{n}\sum_{i=0}^{n}(x_i - \overline{x})(x_i - \overline{x})^T \tag{1}$$

where $\overline{x}$ denotes the average of x_i.

2.2 Classical LDA

LDA is a well-known technique for finding a set of projecting vector W_{LDA} best discriminating different classes. The within-class scatter matrix S_w and the between-class scatter matrix S_b are defined as below:

$$S_w = \frac{1}{n}\sum_{j=1}^{c}\sum_{x \epsilon c_j}(x - \overline{x}_j)(x - \overline{x}_j)^T \tag{2}$$

$$S_b = \frac{1}{n}\sum_{j=1}^{c} n_i(\overline{x}_j - \overline{x})(\overline{x}_j - \overline{x})^T \tag{3}$$

where n_i denotes the number of samples in class $c_j(j - 1, 2, ...c)$, and n denotes the total number of samples. $\overline{x}_j$ denotes the average of samples in c_j^{th} class, x denotes the average of all training samples. One way to find the transformation matrix W_{LDA} is to use Fisher's criterion. It can be achieved by maximizing the ratio as shown in below equation:

$$J(W_{LDA}) = \frac{W_{LDA}^T S_b W_{LDA}}{W_{LDA}^T S_w W_{LDA}} \tag{4}$$

W_{LDA} can be constructed by the set of largest eigenvalues of $S_w^{-1}S_b$. The maximum value of discriminative space is $c - 1$, where c denote the number of classes, since the rank of S_b is $c - 1$. The number of dimensions after the dimension reduction step is less than 10 for 11 classes presented in this paper.

2.3 Relevance weighted LDA(RW-LDA)

The classical LDA criterion is not optimal with respect to minimizing the classification error rate in the lower dimensional space. He tends to overemphasize the classes that are more separable in the input feature space. As a result, the classification ability will be impaired. To deal with this problem, Loog et. al. [2] have proposed to introduce a weighting function to the discriminant criterion, where a weighted between-class scatter matrix is defined to replace the conventional between-class scatter matrix. Classes that are closer to each other in the output space, and thus can potentially impair the classification performance, should be more heavily weighted in the input space. According to [2], weighted between-class scatter matrix $\hat{S}_b$ can be defined as:

$$\hat{S}_b = \sum_{i=1}^{c-1}\sum_{j=i+1}^{c} w(d_{ij})p_i p_j(m_i - m_j)(m_i - m_j)^t \tag{5}$$

Where p_i and p_j are the class priors, d_{ij} is the Euclidean distance between the means of class i and class j. The weighting function $w(d_{ij})$ is generally a monotonically decreasing function:

$$w(d_{ij}) = d_{ij}^{-2h} \text{ with } h \epsilon N \tag{6}$$

Recently, [6] has extended the concept of weighting to estimate a within-class scatter matrix. Thus by introducing a so-called relevance weights,a weighted within-class scatter matrix $\hat{S}_w$ is defined to replace a conventional within-class scatter matrix:

$$\hat{S}_w = \sum_{i=1}^{c-1} p_i r_i \sum_{j=1}^{n_i}(x_{ij} - m_i)(x_{ij} - m_i)^t \tag{7}$$

Where r_i's$(0 < r_i \leq 1, \forall i)$ are the relevance based weights defined as:

$$r_i = \sum_{j \neq 1} \frac{1}{w(d_{ij})} \tag{8}$$

Using the weighted scatter matrices $\hat{S}_b$ and $\hat{S}_w$, the criterion is weighted and the resulting algorithm is referred to as relevance weighted LDA(RW-LDA).

2.4 LDA-QR

Recently, Ye et. al. [4] have proposed a novel algorithm namely LDA/QR. It achieves the efficiency by introducing QR decomposition on a small size matrix. It is also stable since all the decomposition and the inversions are applied to small size matrices. LDA/QR contains two stages. The first one is to maximize separability between different classes. This is done by solving the following optimization criterion:

$$W = arg \max_{W^tW=1} trace(W^t S_b W) \tag{9}$$

The solution to (9) can be obtained through QR decomposition as follows:

Let $H_b = QR$ be the QR decomposition on H_b, where $Q \epsilon R^{dxt}$ has orthonormal columns, $R \epsilon R^{txt}$ is an upper triangular matrix and $t = rank(H_b)$. Then $W = QG$ for any orthogonal matrix $G \epsilon R^{txt}$ solves the optimization problem in (9). The second stage of LDA/QR incorporates the within-class scatter information by applying a relaxation scheme on G. The original problem of finding optimal W is equivalent to finding G, such that

$$G = arg \min_{G} trace((G^t \hat{S}_b G)^{-1}(G^t \hat{S}_w G)) \tag{10}$$

Where $\hat{S}_b = Q^t S_b Q$ and $\hat{S}_w = Q^t S_w Q$ Note that $\hat{S}_b$ and $\hat{S}_w$ have much smaller size than the original scatter matrices S_b and S_w respectively. Now, the optimization problem in [1] can be solved efficiently by solving a small eigen problem on $\hat{S}_b^{-1} \hat{S}_w$.

3 The Proposed Algorithm

We take input data as matrix and reduce the dimensions of the input matrix. After that relevance weights used to discriminate among the classes. When a small sample size (SSS) problem takes place, such as in the face recognition area within class scatter matrix (sw) is also singular so LDA-RW will break down. When small sample size problem is encountered, we propose to use singular value decomposition (SVD) to overcome this problem. We propose to use SVD with modified equations of H_b and H_w defined as follows using LDA-RW method:-

$$H_b = [\alpha_{12}(m_1 - m_2), \ldots \alpha_{(c-1)c}(m_{c-1} - m_c)] \tag{11}$$

$$H_w = [\beta_1(X_1 - m_1 e_1), \ldots.., \beta_c(X_c - e_c m_c)] \tag{12}$$

So that $S_b = H_b H_b^t$and $S_w = H_w H_w^t$

Where $\alpha_{ij} = (W(D_{ij}))^{1/2}$ and $\beta_i = (r_i)^{1/2}$

$w(d_{ij})$ is a weighting function defined as:

$$w(d) = ((m_i - m_j)^\tau (m_i - m_j))^{-h}, h > 0 \tag{13}$$

the steps are followed which presents the pseudo code:-

3.1 Algorithm: 2DLDA/SVD

Input: Data matrix $X, h, l1, l2$

Output: Discriminant projection matrix W

1. Compute the mean M_i of i^{th} class for each i as $M_i = \frac{1}{n} \sum_{x \epsilon \pi_i} X$;
2. Compute the global mean $M = \frac{1}{n} \sum_{i=1}^{k} \sum_{x \epsilon \pi_i} X$;
3. $R_1 \leftarrow (I_{l2}, 0)^T$;
4. For j from 2 to $N+1$
5. $S_w^R \leftarrow \sum_{i=1}^{k} \sum_{x \epsilon \pi_i} (X - M_i) R_{j-1} R_{j-1}^T (X - M_i)^T$,
   $S_b^R \leftarrow \sum_{i=1}^{k} n_i (M_i - M) R_{j-1} R_{j-1}^T (M_i - M)^T$;
6. Compute the first $l1$ eigen vectors of $(S_w^R)^{-1} S_b^R$ and assign it to L_j
7. $S_w^L \leftarrow \sum_{i=1}^{k} \sum_{x \epsilon \pi_i} (X - M_i)^T L_j L_j^T (X - M_i)$,
   $S_b^L \leftarrow \sum_{i=1}^{k} n_i (M_i - M)^T L_j (L_j)^T (M_i - M)$;
8. Compute the first $l2$ eigen vectors of $(S_w^L)^{-1} S_b^L$ and assign it to R_j
9. End For
10. $L \leftarrow L_{(N+1)}, R \leftarrow R_{(N+1)}$;
11. Reduce input data dimension using $X = L^T X R$
12. Construct H_b and H_w
13. Perform SV decomposition: $H_b = USV$
14. Compute $\hat{S}_b = U^t S_b U and \hat{S}_w = U^t S_w U$
15. Compute the t eigenvectors g_i of $\hat{S}_w^{-1} \hat{S}_b$ in increasing order, where $t = rank(H_b)$
16. The projection matrix is $W = UG$
    Where $G = [g_1, g_2, \ldots \ldots, g_t]$

4 Experiment Results

The ORL faces database was used to test the proposed method. In our experiments, we have used a 2.67 GHz Pentium 4 computer with 512 MB RAM and the MATLAB development environment.

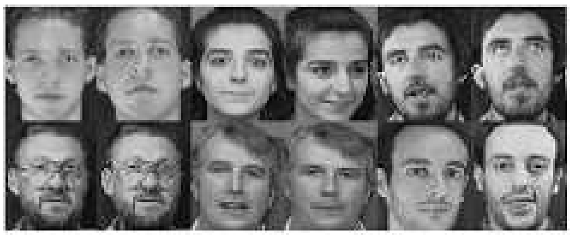

Figure 1: ORL face database: Example images

4.1 Experiments with the ORL Face database

The ORL face database consists of images from $c = 40$ different people, using 10 images from each person, for a total of 400 images. For the test we randomly take k images from each class as the training data, with k in $(2,\ldots,9)$, and leave the rest images to test. Thus we are identifying which image is exactly recognizes in same class or which not. In addition to our proposed method we tested Fisher face method and LDA/QR method also. K represents the number of images in each class.

K	Fisher	LDA/QR	2DLDA/SVD
2	63.75	70.25	77.17
3	71.17	72.5	83.17
4	75	76.17	87.25
5	75.5	78.17	88.75
6	81.87	82.75	89.87
7	82.5	85.17	90.25
8	83.75	86.87	92.75
9	87.5	90	94.5

Table 1: Test results: Recognition rate comparison

4.2 Time Complexity

Test are carried out by well known ORL database with taking c=6 classes and 10 images in each class In graph we have shown comparative execution time complexity of FisherFace, LDA/QR and our proposed algorithm. This Graph shows that our proposed algorithm is much time efficient.

From these results we can see that LDA/SVD gives better results.

5 Conclusion

In this paper, we proposed a novel method for face recognition. This method combines the advantages of the recent LDA enhancements namely relevance weighted LDA and Singular value decomposition and also reduces dimensions of input data matrix using 2DLDA concept. Our method gives better results then previous methods on the bases of face recognition rate and time complexity.

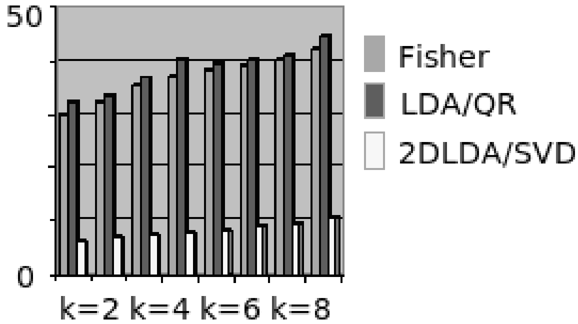

Figure 2: Time complexity graph

References

[1] H. Yu, J. Yang, A direct LDA algorithm for high-dimensional data with application to face recognition, *Pattern Recognition*, 2004, vol. 34, pp. 2067-2070

[2] M. Loog, R. P. W. Duin, R. Hacb-Umbach, Multiclass linear dimension reduction by weighted pairwise Fisher criteria, *Pattern Recognition and Machine Intelligence*, 2001, vol. 23, pp. 762-766

[3] P. Belhumeur J. P. Hespanha and D. J. Kriegman, Eigenfaces vs. Fisherface; Recognition using class specific linear projection, *Pattern Recognition and Machine Intelligence*, 1997, vol. 17(9), pp. 711-720

[4] Jieping Ye, Ravi Janardan, Qi Li, Two-Dimensional Linear Discriminant Analysis, *Advances in neural information processing system*, 2005, vol. 17, pp. 1569-1576

[5] J.Ye, Q Li, LDA/QR: an efficient and effective dimension reduction algorithm and its theoretical foundation, *Pattern Recognition*, 2004, vol. 37(4), pp. 851-854

[6] E. K. Tang, P. N. Suganthan, X. Yao, A. K. Qin, Linear dimensionality reduction using relevance weighted LDA, *Pattern Recognition*, 2005, vol. 38, pp. 485-493

ISBN:978-988-98671-9-5