International Journal of Information Technology and Knowledge Management
July-December 2011, Volume 4, No. 2, pp. 603-607

FACE RECOGNITION USING LDA WITH WAVELET TRANSFORM APPROACH

Neeta Nain[1], Akshay Kumar[2], Amlesh Kumar Mohapatra[3], Ratan Das[4], Ashok Kumar[5] & NemiChand Singh[6]

Linear Discriminant Analysis (LDA) is one of the principal techniques used in face recognition systems. Linear Discriminant Analysis (LDA) is well-known scheme for feature extraction and dimension reduction. It provides improved performance over the standard Principal Component Analysis (PCA) method of face recognition by introducing the concept of classes and distance between classes. This paper provides an overview of PCA, the various variants of LDA and their basic drawbacks. Also has been proposed a development over classical LDA i.e. LDA using wavelets transform approach that enhances performance as regards accuracy and time complexity. Experiments on ORL face database clearly demonstrate this and the graphical comparison of the algorithms clearly showcases the improved recognition rate in case of the proposed algorithm.

Keywords: Face Recognition, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Relevance Weighted LDA, LDA/QR, Wavelet Transform, Sub-bands.

1. Introduction

The vital role played by face recognition systems in today's world has lead to a lot of research being done in this field. 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) and Linear Discriminant Analysis (LDA) are the basic face recognition techniques. 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 projecting 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.

PCA although effective shows decreased efficiency in case of angular lighting and large database. Here LDA comes of use bringing in the concept of within class and between classes variance. 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. 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 neighbour classes also occurs. To overcome this problem an extended scheme is also proposed by applying weighting function in the estimation of scatter matrices. M. Loog [2] has presented an LDA enhancement algorithm namely relevance weighted LDA (RW-LDA) by replacing the un-weighted scatter matrices through the weighted scatter matrices in the classical LDA method. A critical issue using LDA is the small sample size (SSS) problem [3]. 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 the execution of LDA may encounter computational difficulty. In the past, many LDA extensions have been developed to deal with this singularity problem. Among them most popular one is using PCA as a pre-processing step and then performs LDA so dimensionality reduction occurs during PCA phase. But intermediate dimension reduction may lead to information loss. Recently, an efficient variant of LDA, namely LDA/QR, was proposed. The essence of LDA/QR is the utilization of QR-decomposition to reduce dimension of scatter matrix by producing orthonormal matrix during decomposition stage. The time complexity of LDA/QR is linear in the size of the training data, as well as the number of dimensions of the data.

Traditionally, classical LDA is performed on the whole facial image. However, this approach suffers from the limitations mentioned earlier. To resolve these limitations, we propose a new method to use LDA-applying LDA on the wavelet sub-band. The rest of this paper is organized as follows: Section 2 details the existing techniques. Section 3 describes the proposed method in detail. The effectiveness

[1-5]Department of Computer Engineering, MNIT, Jaipur, INDIA
[6]Department of Computer Engineering, GEC, Jhalawar, INDIA
E-mail: [1]neetanain@yahoo. com, [2]akshaykumawat12390@gmail. com, [3]amlesh.mohapatra@gmail.com, [4]rattan.nitj@gmail.com, [5]aashokgahlot@gmail.com, [6]nemi_raj@rediffmail.com

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 (classical LDA) method and the more recent LDA/QR method in Section 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. The PCA and LDA algorithms form the class of "View-Based" face recognition approaches in which the face images are represented as a higher dimensionality vector. The recognition process consists of comparing the test face data in the feature space with the set of feature space data obtained from the training face images. The similarity between these two is then measured to determine if the test face is present in the trained set of images or not.

2.1 Principal Component Analysis (PCA)

Principal Component Analysis is a standard technique used to approximate the original data with lower dimensional feature vectors. The basic approach is to compute the eigenvectors of the covariance matrix and approximate the original data by a linear combination of leading eigenvectors. After subtracting the mean from the individual image pixels stored in matrix form, the covariance matrix(C) is calculated as per the given formula:

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

Where $\overline{x}$ denotes the average of x_i

We then calculate the eigenvectors and eigenvalues of the covariance matrix and choose the feature components from them. The transformation matrix W is defined as the eigenvectors of C corresponding to the K (heuristically chosen) largest eigenvalues.

2.2 Classical LDA

This algorithm emphasizes on discrimination between images by dividing them into classes and clustering the within class objects and distancing them from other class objects. It introduces two matrices: between class scatter matrix (S_B) and within class scatter matrix (S_w). Their computation is as follows:

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

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

Here, n_j denotes the number of samples in class cj(j = 1-.c), and n denotes the total number of samples. x_j denotes the average of samples in class c_j and x denotes the average of all training samples.

We then calculate the eigen vectors using the formula:

$$S_b * W_{lda} = S_w * W_{lda} * D$$

Where W_{lda} is the eigen vector matrix and D is the eigenvalue.

This maximizes the ratio of within class to between class variance and thereby establishes the algorithm. Retain only the C-1 eigenvectors corresponding to the C-1 largest eigenvalues. This gives the basis vector W_{LDA}. Then the image vector X is projected onto this basis vector and the weights of the image are computed.

2.3 PCA LDA Fusions

Face recognition systems have achieved good results but at the expense of robustness. This method aims at improving the robustness of a face recognition system based on the "fusion" of two well-known statistical representations of a face: PCA and LDA. At the same time it is also an improvement over the Small Sample Size (SSS) problem associated with the classical LDA approach. This method incorporates PCA to reduce the overall dimensionality of the feature space. This algorithm selects some of the dimensions of the image and disregards the rest. PCA is used to reduce the dimension of the feature space to (N-C) and then scatter matrices are projected onto this basis vector to obtain non-singular scatter matrices SWN and SBN. The rest of the algorithm proceeds as per the classical LDA steps.

2.4 LDA-QR

Recently, Ye et. al. [3] 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:

$$W = \arg\max_{W_t W = 1} \operatorname{trace}(W_t S_b W) \tag{4}$$

The solution to (4) can be obtained through QR decomposition as follows:

Define matrices

$$H_b = \left[\sqrt{N_1}(m_1 - m)\ldots\sqrt{N_c}(m_c - m)\right] \in IR^{d\times c} \text{ And}$$

$$H_w = \left[\left(X_1 - m_1 e_1^T\right)\ldots\left(X_c - m_c e_c^T\right)\right] \in IR^{d\times n}$$

Where $e_i = (1,\ldots,1)^T \in IR^{n_i}$

Such that $S_b = H_b H_b^T$ and $S_w = H_w H_w^T$

Let H_b = QR be the QR decomposition on H_b, where $Q \in R^{dxt}$ has orthonormal columns, $R \in R^{txt}$ is an upper triangular matrix and t = rank (H_b). Then W = QG for any orthogonal matrix $G \in R^{txt}$ solves the optimization problem in (4). 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

$$G = \arg\min_G \operatorname{trace}\left(\left(G^t S_b G\right)^{-1}\left(G^t S_w G\right)\right) \qquad (5)$$

Where $\hat{S}_b = Q^t S_b Q$ and $\hat{S}_w = Q^t S_w Q$. $\hat{S}_b$ and $\hat{S}_w$ have much smaller size than the original scatter matrices S_b and S_w respectively.

2.5 Relevance Weighted LDA (RW-LDA)

This algorithm is an improvement over the classical LDA algorithm. It successfully separates the different classes and prevents their overlap. Weights are assigned to the scatter matrices by a monastically decreasing weight function as follows:

$$w(d_{ij}) = d_{ij}^{-2h} \ldots \text{with} \ldots h \in N$$

According to [2], weighted between-class scatter matrix $\hat{S}_b$ can be defined as

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

Where p_i and p_j are the d_{ij} class priors and is the Euclidean distance between the means of class i and class j.

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.

3. THE PROPOSED ALGORITHM

To resolve the previous limitations of LDA, we propose an improvement over LDA-applying LDA on the wavelet sub-band. Following sections elaborate our proposed method.

3.1 Wavelet Transform

Wavelet Transform (WT) has been a very popular tool for image analysis in the recent years. The advantages of WT, such as good time and frequency localizations, have been utilised in various areas like image compression, image processing, computer graphics etc. In the proposed system, WT is chosen to be used in image decomposition because:

- By decomposing an image using WT, the resolutions of the sub-band images are reduced. In turn, the computational complexity will be reduced dramatically by working on a lower resolution image.
- Wavelet decomposition provides local information in both space domain and frequency domain.

A two dimensional wavelet transform is derived from two one-dimensional wavelet transform by taking tensor products. The implementation of WT is carried out by applying an one-dimensional transform to the rows of the original image data and the columns of the row transformed data respectively. An image is decomposed into four subbands as shown in Figure 1(b). The band LL is a coarser approximation to the original image. The bands LH and HL record the contours/edges along horizontal and vertical directions respectively. While the HH band records the diagonal edges of the image. This is the first level decomposition. Further decomposition can be conducted on the LL subband.

Fig. 1(a): Face Image From

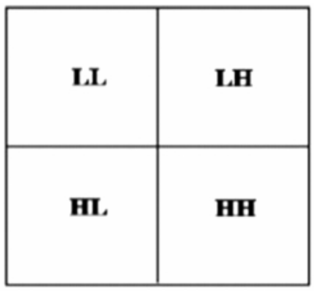

Fig.1(b): Face Image After 1st Level Decomposition ORL Database

3.2 Choosing WT Transform and Subband

In this paper, haar wavelet is adopted for image decomposition. In fact, in order to select a suitable wavelet, the recognition rates are computed by applying different wavelets on face image decomposition. The application of different wavelet transform in image decomposition like haar, Daubechies wavelet, Biortoghonal wavelet, Symlets do not cause major changes in recognition rates. Therefore Haar wavelet is adopted for image decomposition in our system.

In choosing the WT subband, we have the following criteria:

1. Reduce the computational complexity

   To minimize the computational complexity, we choose to work on those subbands with lowest resolution. Accordingly, subband 1 is chosen.

2. Subband 1 contains most of Variance/energy of the original image after applying the wavelet transform.

3.3 Algorithm

1. Apply haar wavelet transform on training face images and get the Subband 1(approximation coefficient band). Now we store the subband images in form of column vectors to form the vector X

   $$X = [x_1, x_2, x_3, \ldots x_n]$$

   Here n is total number of images and $x_1, x_2, x_3 \ldots x_n$ are the subband images corresponding to face image X_i.

2. Calculate global mean as well as the mean of each class C_i (i = 1...c).

3. Calculate the within-class scatter matrix S_w and the between-class scatter matrix S_b as defined in (2), (3).

4. The transformation matrix W is given by the eigenvectors corresponding to the C-1 largest eigenvalues of the matrix $S_w^{-1} S_b$. Vector X is projected onto these eigenvectors and the weights of the images are computed.

5. For a test face image to be recognized, initially haar wavelet is applied to get subband 1 image and then global mean is subtracted from this image. Then the subtracted subband image is projected on the eigenvectors to give the weights of the test image.

6. Compare the weight obtained with the values of weights recorded from the training phase. This comparison is performed using distance metrics such as the Euclidean distance, the L2 or the Mahalanobis distance.

   In this paper, we have used the Euclidean distance to determine the variation of the test image weight from the training set weights. Based on the Euclidean distance the training face with the nearest weight is determined as the best match for identification.

4. EXPERIMENT RESULTS

In our experiments, we have used a 2 GHz computer with 3 GB RAM and the MATLAB development environment. The ORL faces database was used to test the proposed method.

Fig. 2: ORL Database

4.1 Experiments with the ORL Face Database

Tests are carried out by well known ORL database with taking c =10 classes and 10 images in each class For the test we randomly take k images from each class as the training data, with k in (5...9), and leave the rest images to test. Thus we are identifying which image is exactly recognized in same class or which not. In addition to our proposed method we also tested LDA-RW and LDA/QR method also. K represents the number of images in each class.

4.2 Time Complexity

In graph we have shown comparative execution time complexity of Fisherfaces (Classical LDA), LDA-RW, LDA/QR and the proposed approach

Table 1
Test Results: Recognition Rate Comparison

K	LDA	LDA-RW	LDA/QR	Wavelet LDA
5	79	83	87	83
6	85	85	89	83
7	89	90	90	93
8	91	93	94	94
9	97	97	98	95

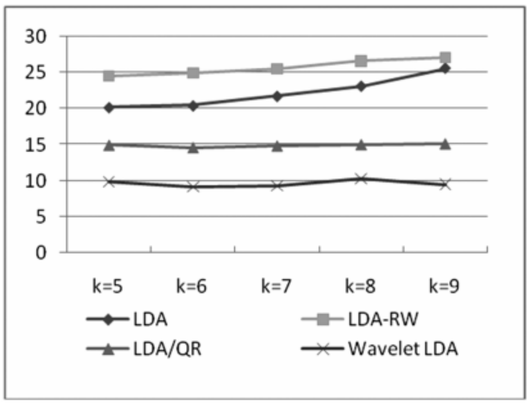

Fig.3: Time Complexity Graph

5. CONCLUSION

In this paper, we proposed a novel method for face recognition. This method combines the wavelet transform and LDA face recognition algorithm Our method gives better results than previous methods as regards face recognition rate and time complexity.

REFERENCES

[1] M.A. Turk and A.P. Pentland. "Face Recognition Using Eigenfaces", In Proc. of Computer Vision and Pattern Recognition, pages 586-591. IEEE, June 1991b.

[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, 23, pp. 762-766.

[3] 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.

[4] Tao Xiong, Jieping Ye. ,Qi Li, "Efficient Kernel Discriminant Analysis via QR Decomposition".

[5] G C Feng , P C Yuen and D Q Dai, "Human Face Recognition Using PCA on Wavelet Subband".

[6] Juwei Lu, Kostantinos N. Plataniotis, "Face Recognition Using LDA-Based Algorithms".