International Journal of Innovative Research in Electronics and Communications (IJIREC)
Volume 2, Issue 2, April 2015, PP 17-22
ISSN 2349-4042 (Print) & ISSN 2349-4050 (Online)
www.arcjournals.org

Performance Analysis of PCA and LDA

K.Raju[1], Dr.Y.Srinivasa Rao[2], Dr.M.Narsing Yadav[3]

[1]Associate Professor,Dept of ECE, KKR & KSR Institute of Technology and Science, Guntur, AP
[2]Professor, Dept. of Instrument Technology, College of Engineering AU, Vizag. AP.
[3]Professor, HOD, Dept of ECE, Malla Reddy Institute of Engineering and Technology, Hyderabad, AP

Abstract: *The problem of face recognition is a composite task that involves the detection and location of faces in a cluttered background, normalization, recognition and verification. Depending on the nature of the application like the sizes of the training and testing databases, clutter and visibility of the background, noise, occlusion and speed requirements, some of these tasks could be very challenging. There have been several methods proposed for face recognition. And one of the key components of any methods is facial feature extraction. There are two major approaches to facial feature extraction for recognition, holistic template matching based systems and geometrical local feature based systems. In this paper we present the methods of PCA and LDA which are based on the holistic approach. The paper first focuses the major aspects of the implementation of both PCA and LDA and finally compare the performance analysis of both the techniques on the standard Image data base of AT&T*

Keywords: *PCA,LDA, holistic, Geometrical local feature based, normalization, recognition, verification*

1. Introduction

Within the last few years, there have been numerous algorithms been proposed for face recognition[2]. While much progress has been made toward recognizing faces under small variations in lighting, facial expression and pose, reliable techniques for recognition under more extreme variations have proven elusive. The detection of a face or a group of faces in a single image or a sequence of images, which has applications in face recognition as well as video conferencing systems, is a challenging task and has been studied by many researchers [4]. Once the face image is extracted from the scene, its gray level and size are usually normalized before storing or testing. In some applications, such as identification of passport pictures or drivers' licenses, conditions of image acquisition are usually so controlled that some of the preprocessing stages may not be necessary.

One of the most important components of an AFR system is the extraction of facial features, in which attempts are made to find the most appropriate representation of face images for identification purposes. The main challenge in feature extraction is to represent the input data in a low-dimensional feature space, in which points corresponding to different poses of the same subject are close to each other and far from points corresponding to instances of other subjects' faces. However, there is a lot of within-class variation that is due to differing facial expressions, head orientations, lighting conditions, etc., which makes the task more complex.

Closely tied to the task of feature extraction is the intelligent and sensible definition of similarity between test and known patterns. The task of finding a relevant distance measure in the selected feature space, and thereby effectively utilizing the embedded information to identify human subjects accurately, is one of the main challenges in face identification. In this paper we focus on feature extraction and face-identification processes using the two commonly used techniques of PCA and LDA and bring out the comparative analysis between the two techniques.

Both PCA and LDA are used for data classification and dimensionality reduction. Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their performances has been examined on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability. The use of Linear Discriminant Analysis for data classification is applied to classification problem in speech recognition. In this paper we show that the results obtained by LDA are better in classification compared to Principal Components Analysis (PCA). The prime difference between LDA and PCA is that PCA does more of feature classification and LDA does data classification. In PCA, the shape and location of the original data sets changes when transformed to a different space whereas LDA doesn't change the location but only tries to provide more class separability and draw a decision region between the given classes. This method also helps to better understand the distribution of the feature data.

2. Principal Component Analysis (PCA)

Before PCA, probably the simplest classifier was using the correlation method, however, they are computationally expensive and would require great amounts of storage. Hence a better technique, which could pursue the dimensionality reduction was proposed. One such technique which is most commonly used for dimensionality reduction in computer vision particularly in face recognition is the PCA [6]. PCA techniques also known as Karhunen-Loeve methods choose a dimensionality reducing linear projection that maximizes the scatter of all the projected samples. PCA is a simple, non-parametric method extracting relevant information from confused data sets. Application of PCA in face recognition linearly projects the image space to a low dimensional feature space, which gives projection directions that maximize the total scatter across all classes, i.e. across all images of all faces.

To more clearly illustrate the above principle, let us consider the analysis with the help of an image terminology.

With X_i representing the image vector of the i-th image and

$$m = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (1)$$

$$X = [X_1 - m, X_2 - m, ..., X_N - m] \qquad (2)$$

The total variance matrix is defined as

$$C = XX^T = \frac{1}{N}\sum_i \sum_i \left(\overline{x_{ij}} - \overline{m}\right)\left(\overline{x_{ij}} - \overline{m}\right)^T \qquad (3)$$

The goal is to obtain N eigen vectors of C such that $C\vec{w} = \lambda \vec{w}$ if we directly perform the eigen –decomposition on C, it is computationally too tedious. Hence as C is highly over determined, the eigen decomposition is performed on $X^T X$instead of XX^T.

$$X^T X\vec{u} = \lambda\vec{u}$$

$$CX\vec{u} = \lambda X\vec{u}$$

$$\therefore \vec{w} = X\vec{u} \qquad (4)$$

The new feature vectors, $y_k \in \Re^m$ are defined by the following linear transformation.

$$y_k = W^T x_k \quad k = 1, 2, ..., N \ , \qquad (5)$$

Where $W \in \Re^{n \times m}$is a matrix with orthonormal columns. Once the linear transformation W^T is applied the variance of the transformed feature vector $\{y_1, y_2, ..., y_N\}$ is $W^T CW$. In this process, the set of eigen vector with eigen value greater than 1 are retained and form the matrix W_p. The projection of all the training images on W_p is

$$y_i = W_p^T (X_i - m) \qquad (6)$$

The projection of each test image onto the PCA basis is given as

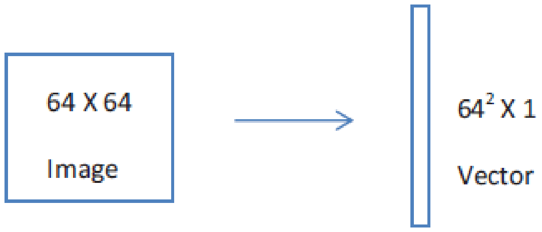
$$y_i' = W_p^T (X_i' - m) \qquad (7)$$

Finally the Euclidean distance between the two projections is calculated and the nearest neighbor in the distance space will be recognized as the matched image. The accuracy of calculation is defined as

$$\text{accuracy(p)} = \frac{Number\ of\ images\ correctly\ recognized}{Total\ number\ of\ images} \tag{8}$$

2.1. PCA Procedure:

Vectorize face images in training and testing database (make column vectors out of each of them), and subtract the mean value of the image from each image vector.



Normalizing the vector to make the norm of the image vector equal to 1, which add tolerance of difference illumination among images.

Training Steps:

1. Calculate the mean of the input face images

2. Subtract the mean from the input images to obtain the mean-shifted images

3. Compute covariance matrix for the image vectors. Calculate the eigenvectors and eigen values of the mean-shifted images

4. Order the eigenvectors by their corresponding eigenvalues, in decreasing order

5. Retain only the eigenvectors with the largest eigenvalues. (the principal components)

    a. The remaining feature set could be with eignevalues greater than 1. If a given eigenvalue is greater than 1, the vectors are stretched in the direction of the corresponding eigenvector.

6. Project the mean-shifted images into the eigenspace using the retained eigenvectors

Classification (Testing) Steps

After the face images have been projected into the eigenspace, the similarity between any pair of face images can be calculated by finding the Euclidean distance between their corresponding feature vectors

Classification is performed using a nearest neighbor classifier in the reduced feature space. The similarity score is calculated between an input face image and each of the training images.

Though the PCA method is relatively simple among the two, a drawback of this approach is that the scatter being maximized is not due to the between class scatter that is useful for classification but also the within class scatter that for classification purposes is unwanted information. As was mentioned by Moses, Adini and Ullman[12], much of the variation from one image to the next is due to illumination changes. Hence is PCA is presented with images of face under varying illumination, the projection matrix will contain the principal components which retain, in the projected feature space, the variation due to lighting. Consequently, the points in the projected space will not be clustered and the worse the classes may be smeared together.

3. Linear Discriminant Analysis (LDA)

As highly structured two-dimensional patterns, human face images can be analyzed in the spatial and the frequency domains. These patterns are composed of components that are easily recognized at high levels but are loosely defined at low levels of our visual system [10].

Each of the facial components (features) has a different discrimination power for identifying a person or the person’s gender, race, and age. There have been many studies of the significance of such features that used subjective psychovisual experiments [11]

In understanding LDA, let us consider the linear transformation since linear discriminant analysis frequently achieves good performances in the tasks of face and object recognition, even though the assumptions of common covariance matrix among groups and normality are often violated (Duda, et al., 2001). In addition, kernel tricks can be used with linear discriminant analysis for non-linear transformation (Mika, et al., 1999). The basic idea of LDA is to find a linear transformation that best discriminate among classes and the classification is then performed in the transformed space based on some metric such as Euclidean distance. Mathematically a typical LDA implementation is carried out via scatter matrix analysis (Fukunaga, 1990).

3.1. Two-class LDA

Fisher first introduced LDA for two classes and its idea was to transform the multivariate observations x to univariate observations y such that the y's derived from the two classes were separated as much as possible. Suppose that we have a set of m p-dimensional samples x1, x2, · · · , xm (where xi = (xi1, · · · , xip)) belonging to two different classes, namely c1 and c2. For the two classes, the scatter matrices are given as

$$S_i = \sum_{\mathbf{x} \in c_i} (\mathbf{x} - \bar{\mathbf{x}}_i)(\mathbf{x} - \bar{\mathbf{x}}_i)', \tag{8}$$

Where $\bar{\mathbf{x}}_i = \frac{1}{m_i} \sum_{\mathbf{x} \in c_i} \mathbf{x}$ and m_i is the number of samples in C_i. Hence the total intra class scatter matrix is given by

$$\hat{\Sigma}_w = S_1 + S_2 = \sum_i \sum_{\mathbf{x} \in c_i} (\mathbf{x} - \bar{\mathbf{x}}_i)(\mathbf{x} - \bar{\mathbf{x}}_i)'. \tag{9}$$

The inter class scatter matrix is given by

$$\hat{\Sigma}_b = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'. \tag{10}$$

Fisher's criterion suggests that the linear transformation Φ to maximize the so called Rayleigh coefficient, i.e., the ratio of the determinant of the inter class scatter matrix of the projected samples to the intra class scatter matrix of the projected samples.

$$\mathcal{J}(\Phi) = \frac{|\Phi^T \hat{\Sigma}_b \Phi|}{|\Phi^T \hat{\Sigma}_w \Phi|}. \tag{11}$$

3.2. Multi Class LDA

If the number of classes are more than two, then a natural extension of Fisher Linear discriminant exists using multiple discriminant analysis. As in two-class case, the projection is from high dimensional space to a low dimensional space and the transformation suggested still maximize the ratio of intra-class scatter to the inter-class scatter. But unlike the two-class case, the maximization should be done among several competing classes. Suppose that now there are n classes. The intra-class matrix is calculated similar to Equation (9):

$$\hat{\Sigma}_w = S_1 + \cdots + S_n = \sum_{i=1}^{n} \sum_{\mathbf{x} \in c_i} (\mathbf{x} - \bar{\mathbf{x}}_i)(\mathbf{x} - \bar{\mathbf{x}}_i)'. \tag{12}$$

The inter class scatter matrix slightly differs in computation and is given by

$$\hat{\Sigma}_b = \sum_{i=1}^{n} m_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \tag{13}$$

Where m_i is the number of training samples for each class and $\bar{\mathbf{x}}$ is the total mean vector given by

$$\bar{\mathbf{x}} = \frac{1}{m} \sum_{i=1}^{n} m_i \bar{\mathbf{x}}_i \tag{14}$$

After obtaining $\hat{\Sigma}_b$ and $\hat{\Sigma}_w$ the linear transformation Φ. It can be shown that the transformation Φ can be obtained by solving the generalized eigen value problem

$$\hat{\Sigma}_b \Phi = \lambda \hat{\Sigma}_w \Phi \tag{15}$$

We can easily deduce the upper and lower bounds of the rank of $\hat{\Sigma}_w$ and $\hat{\Sigma}_b$ are m-n and n-1 respectively. Multiple discriminant analysis provides an elegant way for classification using discriminant features.

3.3. Classification

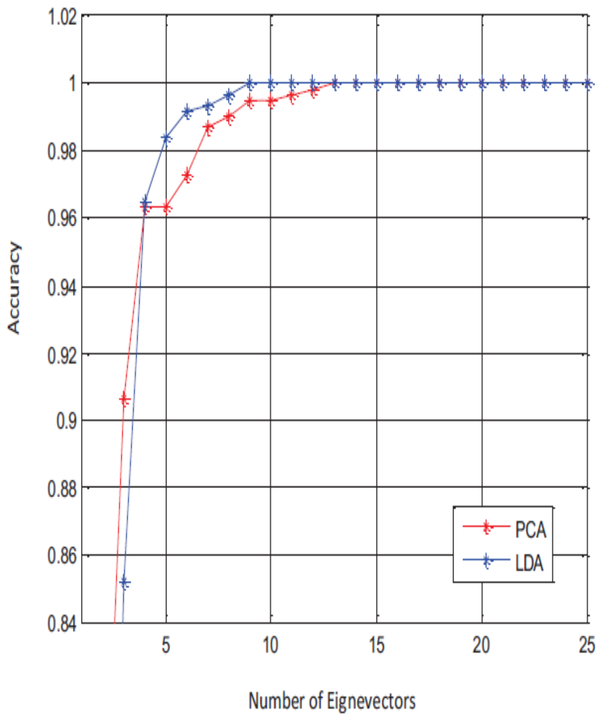
Once the transformation Φ is found, the classification is then performed in the transformed spacebased on some distance metric, such as Euclidean distance $d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_i (x_i - y_i)^2}$ and cosine measure $d(\mathbf{x}, \mathbf{y}) = 1 - \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}}$. Then upon arrival of the new instance z, it is classified to

$$\arg\min_k d(\mathbf{z}\Phi, \bar{\mathbf{x}}_k \Phi) \tag{16}$$

Where $\bar{\mathbf{x}}_k$ is the centroid of the Kth class.

4. Comparison of PCA and LDA

The following figure shows a comparison of PCA and LDA results. LDA achieve 100% accuracy rate after using 9 eigenvectors as feature space. PCA achieve 100% accuracy after using 13 eigenvectors. The PCA projections are optimal for representation in a low dimensional basis, but they may not be optional from a discrimination standpoint. The LDA maximizes the ratio of between-class variance to that of within-class variance. It works better than PCA for purpose of discrimination. In sum, LDA outperformed the PCA method. Both has good performance under varying illumination.



References

[1]. R. Baron, ''Mechanisms of human facial recognition,'' Int. J.Man–Machine Studies **15**, 137–178 (1981).

[2]. G. Davies, H. Ellis, and E. J. Shepherd, *Perceiving and Remembering Faces* Academic, New York, 1981).

[3]. Y. Yacoob and L. S. Davis, ''Computing spatio-temporal representations of human faces,'' in *Proceedings of the IEEEComputer Society Conference on Computer Vision and PatternRecognition* (IEEE Computer Society, Los Alamitos, Calif., 1994), pp. 70–75.

[4]. R. Brunelli and T. Poggio, ‘‘HyperBF networks for gender classification,’’ in *Proceedings of the DARPA Image UnderstandingWorkshop* (Defense Advanced Research Projects Agency, Arlington, Va., 1992), pp. 311–314.

[5]. B. A. Golomb and T. J. Sejnowski, ‘‘SEXNET: A neural network identifies sex from human faces,’’ in *Advances in NeuralInformation Processing Systems 3*, D. S. Touretzky and R. Lipmann, eds. (Morgan Kaufmann, San Mateo, Calif., 1991), pp. 572–577.

[6]. A. Samal and P. Iyengar, ‘‘Automatic recognition and analysis of human faces and facial expressions: a survey,’’ Pattern Recog. **25**, 65–77 (1992).

[7]. R. Chellappa, C. L. Wilson, and S. Sirohey ‘‘Human and machine recognition of faces, a survey,’’ Proc. IEEE **83**, 705–740 (1995).

[8]. V. Govindaraju, S. N. Srihari, and D. B. Sher, ‘‘A computational model for face location,’’ in *Proceedings of the ThirdInternational Conference on Computer Vision* (IEEE Computer Society Press, Los Alamitos, Calif., 1990), pp. 718–721.

[9]. A. Shio and J. Sklansky, ‘‘Segmentation of people in motion,’’ in *Proceedings of the IEEE Workshop on Visual Motion* (Institute of Electrical and Electronics Engineers, Piscataway, N.J., 1991), pp. 325–332.

[10]. G. Yang and T. S. Huang, ‘‘Human face detection in a scene,’’ in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Institute of Electrical and Electronics Engineers, Piscataway, N.J., 1993), pp. 453–458.

[11]. A. Pentland and B. Moghaddam, ‘‘Probabilistic visual learning for object detection,’’ in *Proceedings of the International Conference on Computer Vision* (IEEE Computer Society Press, Los Alamitos, Calif., 1995), pp. 786–793.

[12]. K. Sung and T. Poggio, ‘‘Example-based learning for view based human face detection,’’ *Proceedings of the IEEE Im*-Fig. 10. Cluster separation in the best two-dimensional feature space. Top, based on LDA; bottom, based on PCA. 1732 J. Opt. Soc. Am. A/ Vol. 14, No. 8/August 1997 K. Etemad and R. Chellappa*age Understanding Workshop* (Institute of Electrical and Electronics Engineers, Piscataway, N.J., 1994), pp. 843–850.

[13]. C. Wu and J. Huang, ‘‘Huang face profile recognition by computer,’’ Pattern Recog. **23**, 255–259 (1990).

[14]. G. Gordon, ‘‘Face recognition based on depth maps and surfacecurvature,’’ in *Geometric Methods in Computer Vision*, B. C. Vemuri, ed., Proc. SPIE **1570**, 234–247 (1991).

[15]. O. Nakamura, S. Mathur, and T. Minami, ‘‘Identification of human faces based on isodensity maps,’’ Pattern Recog. **24**,263–272 (1991).

[16]. Y. Cheng, K. Liu, J. Yang, and H. Wang, ‘‘A robust algebraic method for human face recognition,’’ in *Proceedings of the 11th International Conference on Pattern Recognition*

[17]. (IEEE Computer Society Press, Los Alamitos, Calif., 1992), pp. 221–224.

[18]. (IEEE Computer Society, Los Alamitos, Calif., 1991), pp.586–591.

[19]. A. Pentland, B. Moghaddam, T. Starner, and M. Turk, ‘‘View-based and modular eigenspaces for face recognition,’’