Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) based Face Recognition

Shailaja A. Patil
Department of Electronics & Telecommunication Engineering,
R. C. Patel Institute of Technology,
Shirpur, Maharashtra, India.

Pramod J. Deore, Ph.D
Department of Electronics & Telecommunication Engineering,
R. C. Patel Institute of Technology,
Shirpur, Maharashtra, India.

ABSTRACT

This paper presents study of face recognition system which is based on Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) [1], [2]. These methods are used for feature extraction and dimension reduction. Nearest Neighbour Classifier (NNC) is used for classification. For matching Mahalanobis Cosine (Mahacos) and Cosine (Cos) distance is used.

Keywords:

Face Recognition, PCA, LDA, NNC, Cos, Mahacos.

1. INTRODUCTION

Face recognition is a visual pattern recognition problem. There, a face as a three-dimensional object subject to varying illumination, pose, expression and so on is to be identified based on its two-dimensional image (three-dimensional images e.g. obtained from laser may also be used).

A face recognition system generally consists of four modules: detection, alignment, feature extraction, and matching, where localization and normalization (face detection and alignment) are processing steps before face recognition (facial feature extraction and matching) is performed.

Face detection segments the face areas from the background. In the case of video, the detected faces may need to be tracked using a face tracking component. Face alignment is aimed at achieving more accurate localization and at normalizing faces thereby whereas face detection provides coarse estimates of the location and scale of each detected face. Facial components, such as eyes, nose, and mouth and facial outline, are located; based on the location points, the input face image is normalized with respect to geometrical properties, such as size and pose, using geometrical transforms or morphing. The face is usually further normalized with respect to photo-metrical properties such illumination and gray scale.

After a face is normalized geometrically and photometrically, feature extraction is performed to provide effective information that is useful for distinguishing between faces of different persons and stable with respect to the geometrical and photo-metrical variations. For face matching, the extracted feature vector of the input face is matched against those of enrolled faces in the database; it outputs the identity of the face when a match is found with sufficient confidence or indicates an unknown face otherwise. Face recognition results depend highly on features that are extracted to represent the face pattern and classification methods used to distinguish between faces whereas face localization and normalization are the basis for extracting effective features.

2. LITERATURE REVIEW

The current face recognition system algorithms can be classified into two classes, image template based and geometry feature-based image. The template based methods [4] measure the correlation between a face image and one or more model templates to estimate the face image identity. In some papers the statistical tools such as Support Vector Machines (SVM) [6], [7], Linear Discriminant Analysis (LDA) [10], Principal Component Analysis (PCA) [8], [9], kernel methods [11], [12], and neural networks [13], [14], [15] have been used to construct a suitable set of face template images. Other than statistical analysis and neural network approach there are many other approaches known as hybrid approaches in which there is a combination of both statistical pattern recognition techniques and neural network systems. Poggio and Brunelli [5] explain that the optimal strategy for face recognition system is holistic and corresponds to image template matching. In this study, they compared their results with respect to a geometric feature based technique with a template matching based system and reported an accuracy of 90% for the first one and 100% for the second one on a database of 97 persons. In the some examples of hybrid approaches include the combination of Radial Basis Function (RBF) neural network and PCA [16], [17]. Among other methods, people have used range [3], profile [19] and infra-red scanned [18] images for face recognition. While templates can be viewed as features, they mostly capture global features of the face image. 3-dimensional Facial occlusion and illumination variation is often difficult to handle in these approaches.

The geometry feature based methods analyse explicit local facial features, and their geometric relationships. Cootes et al. [20] have presented an active shape model in extending the approach by Yuille [21]. Wiskott et al. [22] developed an elastic bunch graph matching algorithm for face recognition. Penev et al. [23] developed PCA into Local Feature Analysis (LFA).

3. FACE RECOGNITION ALGORITHM

3.1 Principal Component Analysis

Principal component analysis (PCA) method used for extracting global features from high-dimensional data set. It can also be used to identify patterns in data, and expressing the data in such a way as to highlight their similarities and differences.
Principal Component Analysis (PCA) [8], [9], is a statistical dimensionality-reduction method, which produces the linear least-squares subspace of a training set. The s-dimensional vector representation of each face image in a training set of M images, the PCA could find a t-dimensional subspace whose basis vectors correspond to the maximum variance directions in the original image space. New t-dimensional subspace is normally lower dimensional space than the original space (t <<s). All images of a training set are projected onto the subspace, and then sets of weights W^{train}, which describe the contribution of each vector, are calculated.
To recognize an identity of a test image, the image is projected onto the same subspace, then a set of weights W^{test} is figured out. After the distance measures between W^{test} and each set of W^{train}, the identity of W^{train} is determined to that of a minimum-distanced set in W^{train}. The PCA basis vectors are defined as eigenvectors of the scatter matrix S_T defined as:

$$S_T = \sum_{i=1}^{M} (X_i - \mu) \cdot (X_i - \mu)^T \tag{1}$$

where μ is the mean of all images in the training set and x_i is the i_{th} image in the training set.

3.2 Linear Discriminant Analysis

In pattern recognition Linear Discriminant is a “classical” technique of face recognition system. LDA is used to find a linear combination of features vectors which separate two or more classes of objects or events of data sets. The resulting combination is used as a linear classifier and for dimensionality reduction.
From the new dimensions which is a linear combination of pixel values, form a template. The linear combinations obtained using principal component analysis are called eigenfaces and those obtained from Fisher’s linear discriminant are called Fisher faces.
Linear Discriminant Analysis (LDA) [10] searches the vectors in the underlying space that best discriminate among classes. For all samples of all classes of face images, we define between-class scatter matrix S_b and the within-class scatter matrix S_w.

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (X_i^j - \mu_j) \cdot (X_i^j - \mu_j)^T \tag{2}$$

$$S_b = \sum_{j=1}^{c} (\mu_j - \mu) \cdot (\mu_j - \mu)^T \tag{3}$$

where X_i^j is the i_{th} sample of class j, μ_j is the mean of class j, c is the number of classes, N_j the number of samples in class j, and μ is the mean of all classes. S_w represents the scatter of features around the mean of each face class and S_b represents the scatter of features around the overall mean for all face classes.
The goal is to maximize the value of S_b while minimizing the value of S_w.
The main difference between LDA and PCA method is that PCA is used for feature classification and LDA is used for data classification of given face database. In PCA, the location and shape of the original data sets of face images are changes when transformed to a different space whereas LDA does not change the location of the data sets of face images but only tries to provide more class separability and draw to plot a decision region between the given classes of face images.
The LDA is also called as Fisher’s Discriminant Analysis. The main purpose of the Linear Discriminant Analysis (LDA) is used to find a proper way to represent the face vector space of the original data sets of face images. PCA constructs the face space using the whole face training data as a whole, and not using the face class information and for matching score calculation the Mahalanobis Cosine distance is used. On the other hand, LDA uses class specific information which best discriminates among classes. LDA represents an optimal linear discriminant function which maps the input into the classification space in which the class identification of this sample is decided to based on some metric such as Mahalanobis Cosine and Cosine distance.

3.3 Nearest Neighbour Classifier

The nearest neighbour classier is used to compare the feature vector of the prototype image and the feature vectors stored image in the database. It is obtained by finding the Nearest distance between the prototype image and the database images. Let $C_1, C_2, C_3 ......C_k$ be the k clusters in the ORL database. The class is found by measuring the distance $d(X^{(q)}, C_k)$ between $X^{(q)}$ and the k^{th} cluster C_k. The feature vector with minimum difference is found to be the closest matching vector of database. It is given by

$$d(X^{(q)}, C_k) = Min\{||X^{(q)} - X|| : X \in C_k\} \tag{4}$$

In this paper of face recognition we use two distance measures: Mahalanobis Cosine distance and Cosine distance [1], [2]. Mahalanobis Cosine space is defined as a space where the sample variance of data image along each dimension is one. Therefore, the transformation of a feature vector from face image space to feature space of face image is performed by dividing each coefficient in the vector by its corresponding standard deviation. This transformation of a feature vector then yields a dimensionless feature space of face image with unit variance in each dimension. If there are two vectors x and y in the unscaled PCA space and corresponding vectors m and n in Mahalanobis space. First, we define $\lambda_i = \sigma_i^2$ where λ_i are the PCA eigenvalues, σ_i^2 is the variance along those dimensions and σ_i is the standard deviation. The relationship between the vectors are then defined as:

$$m_i = \frac{x_i}{\sigma_i} \; n_i = \frac{y_i}{\sigma_i} \tag{5}$$

$$d(x, y) = \sqrt{\sum_{i=1}^{k} (m_i - n_i)} \tag{6}$$

Where λ_i is the i^{th} Eigenvalue corresponding to the i^{th} Eigenvector.
Cosine distance measure is defined as follows:

$$d_{cos}(x, y) = -\frac{x.y}{||x||||y||} = -\frac{\Sigma_{i=1}^{k} x_i y_i}{[\Sigma_{i=1}^{k} (x_i)^2 \Sigma_{i=1}^{k} (y_i)^2]^{\frac{1}{2}}} \tag{7}$$

The two vectors x and y are the unscaled LDA space.

4. EXPERIMENT AND RESULTS

In this complete face recognition system the face detection and preprocessing steps are absent. In ORL database the given images are preprocessed images.The experiment was done using PCA and LDA on face images. The images were obtained from ORL database. This database includes the 40 folders containing 10 images in each. The database consist of total 400 images of male and female.
The stepwise process of experiment is as follows:

(1) Load images from a ORL database (database loading).
(2) Partition data into training and test sets. In our case, the first 3 images of each ORL database will serve as the training set and the remaining images will serve as test images (data partitioning).
(3) Compute training and test feature vectors using PCA and LDA method. We use PCA and LDA for feature extraction, and, therefore, first compute the PCA and LDA subspace using the training data from the ORL database. The training data comprises 120 samples (images) with 10304 variables (pixels)(feature extraction).
(4) Compute matching scores between training feature vectors and test/query feature vectors using the Mahalanobis cosine and Cosine distance (matching).
(5) Evaluate results and present performance metrics. Computing ROC (Receiver Operating Characteristics) and CMC (Cumulative Match score) curves.

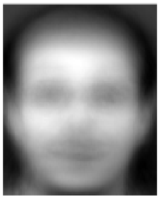

Figure 1. Mean Face (LDA)

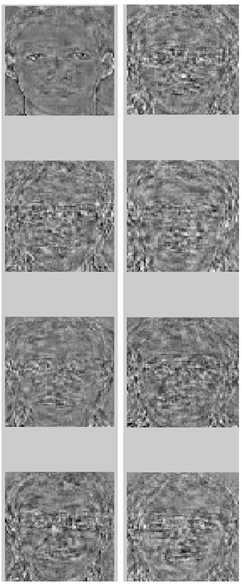

Figure 2. Fisherfaces (LDA)

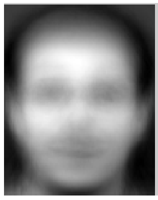

Figure 3. Mean Face (PCA)

Figure 4. Eigenfaces (PCA)

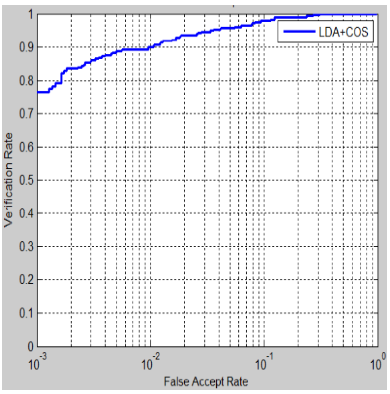

Figure 5. ROC curve for the LDA + Cos technique on the ORL database

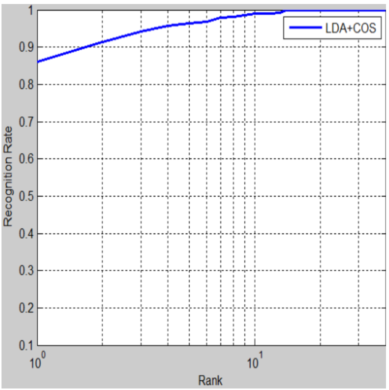

Figure 6. CMC curve for the LDA + Cos technique on the ORL database

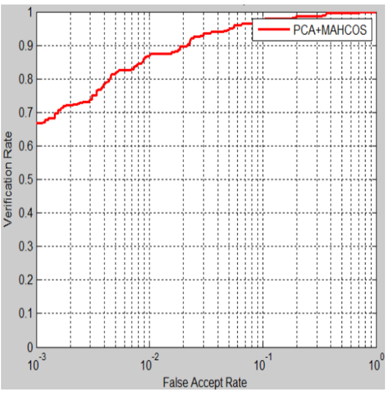

Figure 7. ROC curve for the PCA + Mahcos technique on the ORL database

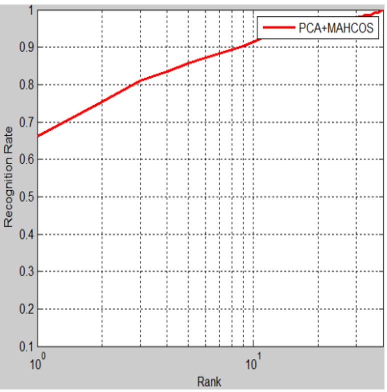

Figure 8. CMC curve for the PCA + Mahcos technique on the ORL database

The Face recognition is done on the basis of indexes which are well accepted to determine the performance of a Face recognition system:

(1) **False Acceptance Rate (FAR)**
Sometimes the biometric security system may incorrectly accept an access attempt of an unauthorized user. To measure these types of incidents FAR is basically used. It measures the percent of invalid inputs which are incorrectly accepted. A systems FAR basically states the ratio between the number of false acceptances and the number of identification attempts.

$$FAR = \frac{FA}{N} \times 100$$

where
FA= number of incidents of false acceptance
N= number of samples

(2) **False Rejection Rate (FRR)**
Sometimes the biometric security system may incorrectly reject an access attempt by an authorized user.To measure these types of incidents FAR is basically used. It measures the percent of valid inputs which are incorrectly rejected.
A systems FRR basically states the ratio between the number of false rejections and the number of identification attempts.

$$FRR = \frac{FR}{N} \times 100$$

where
FR= number of incidents of false rejection
N= number of samples

(3) **Equal Error Rate (ERR)**
The equal error rate (EER) is the rate when false acceptance rate (FAR) is equal to false rejection rate (FRR).
In general, the lower the equal error rate value, the higher the accuracy of the biometric system. However, that most operational systems are not set to operate at the "equal error rate" so the measure true usefulness is limited to comparing biometric system performance. The EER is sometimes referred to as the "Crossover Error Rate".

The results of performance metrics of LDA + Cos and PCA + Mahacos is given in the Table I.

Sr. No.	Performance Metrics	LDA + Cos	PCA + Mahacos
1.	The rank one recognition rate equals (in %)	86.07%	66.07%
2.	The equal error rate equals (in %)	4.28%	5.03%
3.	The minimal half total error rate equals (in %)	4.09%	4.72%
4.	The verification rate at 1% FAR equals (in %)	90.00 %	86.79%
5.	The verification rate at 0.1% FAR equals (in %)	76.43%	66.79%
6.	The verification rate at 0.01% FAR equals (in %)	64.29%	45.00%

Table 1 Comparison of Performance metrics

5. CONCLUSION

In this paper we demonstrates how to use PCA and LDA on real image data and how to classify face images based on PCA, LDA and the nearest neighbour classifier. For matching score calculation the 'Mahcos' and 'Cos' distance is used. After computing the similarity matrix the results are evaluated and some graphical and numerical results are shown.
The plotting function produces a ROC curves that plots the verification rate against the false accept rate and CMC curves that plots the recognition rate against the Rank. we compared the results of performance metrics of PCA + Mahcos and LDA + Cos. From these results here we observed that the rank one recognition rate of LDA + Cos is more than the PCA + Mahcos. The equal error rate

and the minimal half total error rate (in %) of PCA + Mahcos is more than LDA + Cos. The verification rate at 1% FAR, at 0.1% FAR and at 0.01% FAR of PCA + Mahcos is less than LDA + Cos. The results clearly shows that in face recognition system the recognition rate of LDA + Cosine distance performs far better than PCA + Mahalanobis distance based classifier.

6. REFERENCES

[1] Struc V., Paveic, N.,"The Complete Gabor-Fisher Classifier for Robust Face Recognition", EURASIP Advances in Signal Processing, vol. 2010, 26 pages, doi:10.1155/2010/847680, 2010.

[2] Struc V., Paveic, N.,"Gabor-Based Kernel Partial-Least-Squares Discrimination Features for Face Recognition", Informatica (Vilnius), vol. 20, no. 1, pp. 115-138, 2009.

[3] R. Chellappa, C. L. Wilson, and S. Sirohey,"Human and Machine Recognition of Faces: A Survey", Proc. of the IEEE, vol. 83, no. 5, pp. 705-740, May 1995.

[4] Robert J. Baron,"Mechanisms of Human Facial Recognition", International Journal of Man-Machine Studies, vol. 15, no. 2, pp. 137-178, 1981.

[5] R. Brunelli and T. Poggio,"Face Recognition: Features versus Templates", IEEE Tran. on Pattern Analysis and Machine Intelligence, vol. 15, no. 10, pp. 1042-1052, October 1993.

[6] E. Osuna, R. Freund, and F. Girosi,"Training Support Vector Machines: An Application to Face Detection", In IEEE Conference on Computer Vision and Pattern Recognition, pp. 193-199, 1997.

[7] Vladimir N. Vapnik,"The Nature of Statistical Learning Theory", Springer Verlog, Heidelberg, DE, 1995.

[8] L. Sirovich and M. Kirby,"Low-dimensional Procedure for the Characterization of Human Faces", Journal of Optical Society of America, vol. 4, no. 3, pp. 519-524, March 1987.

[9] Matthew Turk and Alex Paul Pentland,"Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

[10] Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman,"Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Tran. on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, July 1997.

[11] Bernhard Scholkopf, Alex J. Smola, and Andre Bernhardt,"Non-linear Component Analysis as a Kernel Eigenvalue Problem", Neural Computation, vol. 10, no. 5, pp. 1299-1319, 1998.

[12] M. H. Yang,"Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition using Kernel Methods", In IEEE International Conference on Face and Gesture Recognition, pp. 215-220, Washington, May 2002.

[13] A. Jonathan Howell and Hilary Buxton,"Invariance in Radial Basis Function Neural Networks in Human Face Classification", Neural Processing Letters, vol. 2, no. 3, pp. 26-30, 1995.

[14] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, and Andrew D. Back,"Face Recognition: A Convectional Neural Network Approach", IEEE Trans. on Neural Networks, vol. 8, no. 1, pp. 98-113, 1998.

[15] T. Poggio and K. K. Sung,"Example-based Learning for View-based Human Face Detection", ARPA Image Understanding Workshop, November 1994.

[16] M. J. Er, S. Wu, and J. Lu,"Face Recognition using Radial Basis Function (RBF) Neural Networks", In 38th Conference on Decision & Control, Phoenix, Arizona USA, pp. 2162-2167, 1999.

[17] C. E. Thomaz, R. Q. Feitosa, and A. Veiga,"Design of Radial Basis Function Network as Classifier in face Recognition using Eigenfaces", In V^{th} Brazilian Symposium on Neural Networks, pp. 118-123, 1998.

[18] Y. Yoshitomi, T. Miyaura, S. Tomito, and S. Kimura,"Face Identification using Thermal Image Processing", In IEEE International Workshop on Robot and Human Communication, pp. 374-379, 1997.

[19] Z. Liposcak and S. Loncaric,"Face Recognition from Profiles using Morphological Operations", In International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, pp. 47-52, 1999.

[20] Andreas Lanitis, Christopher J. Taylor, and Timothy Francis Cootes,"Automatic Interpretation and Coding of Face Images using Flexible Models", IEEE Tran. on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 743-756, 1997.

[21] Alan L. Yuille,"Deformable Templates for Face Recognition", Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 59-70, 1991.

[22] Laurenz Wiskott, Jean-Marc Fellous, Norbert Kruger, and Christoph von der Malsburg,"Face Recognition by Elastic Bunch Graph Matching", IEEE Tran. on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 775-779, July 1997.

[23] P. Penev and J. Atick,"Local Feature Analysis: A General Statistical Theory for Object Representation", Network: Computation in Neural Systems, vol. 7, pp. 477-500, 1996.

[24] Zhao and Pietikinen,"Dynamic texture recognition using local binary patterns with an application to facial expressions", IEEE Trans. Pattern Anal. Mach. Intell. vol. 29, no. 6, pp. 915-928, 2007.

[25] G. Zhao, X. Huang, M. Taini, S. Li, and M. Pietikainen,"Facial expression recognition from near-infra-red videos", Image Vis. Comput, vol. 29, no. 9, pp. 607-619, 2011.

[26] G. Zhao, T. Ahonen, Jiri Matas, and M. Pietikainen,"Rotation-Invariant Image and Video Description With Local Binary Pattern Features", IEEE Trans. Image Process, vol. 21, no. 4, pp. 1465-1477, 2012.