PCA, LDA & Neural Network based Face Recognition

Supriya S.Nalawade, Kshama V. Kulhalli

#-M.E. student of D.Y.Patil college of engg & technology,Kolhapur,Electronics & telecomm Deprtment,Faculty of D.Y.Patil college of engg & technology,Kolhapur,IT Department,Shivaji University
s.s.p.m. college of engg Kankavli,Dist-Sindhudurg,State-Maharashtra,India

[1]supriya32737@yahoo.com

[3]kvkulahalli@gmail.com

A*bstract*— In today's complex, geographically mobile, increasingly electronically inter-connected information society, accurate identification is becoming very important and the problem of identifying a person is becoming ever increasingly difficult. In this fast moving world to keep a watch on every individual is very tiresome and time consuming which is next to impossible.

Automatic face recognition system has been designed to automatically detect the culprits from the data base. Face recognition system deters terrorism, thefts.

***Keywords*—-Pca, Lda and Neural network Back propagation Algorithm**

I. INTRODUCTION

What is biometrics?

Biometrics refers to the automatic recognition of individuals based on their physiological and/or behavioural characteristics. By using Bioinformatics, it is possible to confirm or establish an individual's identity based on "who he is," rather than by "what he possesses" (e.g. an ID card) or "what he remembers" (e.g. a password).

Why we are using face recognition?

Face Recognition is natural and non-intrusive, but it is not trustworthy, while the fingerprint verification is trustworthy but it is intrusive and it can cause resistance to the users depending on the application. And as compared to other biometric systems the cost is very low for face recognition as hardware required is not expensive. Face recognition is a formal method which is first proposed by Francis Galton in 1888. As one of the most successful applications of image.

Analysis and understanding, face recognition has significant demand especially during the past two years.

Challenges to Face Recognition:

The recognition technique in well control environment gives better performance. However, FRT in an uncontrolled environment is still very challenging. There are three major challenges: The Illumination variation problem, Pose variation problem and occlusion. These issues can cause serious performance degradation in FRT.

Illumination: In Illumination problems face appears different due to a change in lighting. Hence performance of FRT becomes poor.

Pose Variance: The performance of Face Recognition System drops significantly when large pose variations are present in the input images. Reducing this difficulty is not difficult, but accurate pose estimation is hard.

Occlusion: Sometimes the face images are partially occluded that is any object comes over the face in the images. And due to this it becomes difficult to recognise a person from an image in the database.

Present State of Technology:

Face recognition is one of the Biometric methods, which identifies individual by features of face. Research in this area has been conducted for more than thirty years as a result; the current status of face recognition technology is well advanced.

1) Face recognition with image sets using Manifold Density Divergence:

This method introduced a new approach to face with image set. It constructs the model that is able to accurately capture the various modes of face appreances under variation in imaging conditions. I this a flexible model is used for learning probability densities confined to highly non-linear but relatively low dimensional manifold. And the divergence between densities estimated on these manifolds can be minimized. It consists of algorithms that are used to match the best, achieving 94% recognition rate.

2) Expression Invariant Face Recognition with expression Classification:

This algorithm utilizes the idea of separating geometry and texture import information in a face image. And then model two types of information by projecting them into separate PCA. And finally the texture and geometry attributes are recombined to form a classifier, which is capable of recognizing faces with different expressions. Since the number of prototypes in each type of expressions is quietly limited and insufficient to cover different ways people defined expressions. The recognition rate is 96%.

3) A Scale Spaced Approach to Face Recognition from profile:

In this method a gray level image of profile is threshold to produce a binary, black and white image. Here, black corresponds to face region. A pre-processing step then extracts outline curve of the front portion of the face that bounds the face image. From this curve a set of twelve fiducial mark is automatically identified. Euclidean distance is used for measuring the similarity of feature vectors that are derived from outline profiles. In this method the obtained recognition rate is 90%.This method is simple and fast.

4) Algorithm Evaluation for Face Recognition: What makes picture difficult.

The first goal of this method is to show that PADO, a variant of genetic programming can produce algorithm that directly recognized the difference between face images. The PADO achieved a 92% recognition rate.

5) Face Recognition using Eigen Faces:

This method recognises the person by comparing characteristics of the face to those of known individual. This approach treats the face recognition as a two dimensional recognition problem taking advantage of the fact that faces are normally up-right and thus may be described by a small set of 2D characteristics views. Face images are projected onto a face space. The face space is defined by "Eigen Faces", which are the Eigen vectors of the set of faces. They do not necessarily correspond to isolated features such as eyes, ears and noses. It is fast and relatively simple and has been shown to work well in a somewhat constraint environment.

6) Face Recognition using LDA based Algorithm:

LDA based algorithms optimizes the low dimensional representation of the object with the focus on most discriminant feature extraction. This method combines the strength of D-LDA and F-LDA approaches while at the same time overcome their shortcoming and limitations.

There are four steps to recognize the face in the proposed method.

1. Pre-processing:

In this part, at first we manually cut the input images to 40× 40 images in order to remove the background information and have only face details (face images in most databases contain background information that is not useful for recognition). After that we histogram equalize all input images in order to spread energy of all pixels inside the image and then normalize them to equalize amount of energy related to each face image. The transformation of the pixel intensity values of the given image via the rank transform. The rank transform is basically histogram equalization procedure which renders the histogram of the given image in such a way that the resulting histogram approximates the uniform distribution.

2. Dimensionality Reduction Using PCA:

Algorithm:

Step 1: Get some data

Step 2: Subtract the mean

For PCA to work properly, you have to subtract the mean from each of the data dimensions. The mean subtracted is the average across each dimension. So, all the x values have NxN (the mean of the x values of all the data points) subtracted, and all the y values have $\bar{y}$ subtracted from them. This produces a data set whose mean is zero.

Step 3: Calculate the covariance matrix

From this new image space of M ϕi images (Each with dimension N x 1), the matrix A is formed with dimension N x M by taking each of image vectors ϕi and placing them in each column of matrix A.

Using matrix A, it is important to set up the Covariance matrix C. This can be given by product of matrix A with matrix A^T . The dimension of such covariance matrix will be N x N.

$C = AA^T$

Step 4: Calculate the eigenvectors and eigenvalues of the covariance matrix

As the dimension of this matrix is N x N, which means it will result in N eigenvalues and N eigenvectors. Since the value of N is very large, say 65536 as in above example, it would be better to reduce this overhead by considering matrix $L = A^TA$. The dimension of this matrix will be M x M.

$L = AA^T$

The N eigenvalues obtained from C are same as M eigenvalues with remaining N - M eigenvalues equals zero. Also if x is eigenvector obtained from C then the eigenvectors of L are given by

$y = A^T x$

We can make use of this relationship to obtain eigenvalues and eigenvectors of AA^T by calculating eigenvalues and eigenvectors for A^TA . The eigenvectors for C (Matrix U) are obtained from eigenvectors of L (Matrix V) as given below:

$$U = AV$$

The matrix V, with dimension (M x M), is constituted by the M eigenvectors of L and matrix U, with dimension (N x M), is constituted by all the eigenvectors of C, and the matrix A is the image space, with dimension (N x M).

Step 5: Choosing components and forming a feature vector

In general, once eigenvectors are found from the covariance matrix, the next step is to order them by eigenvalue, highest to lowest. This gives you the components in order of significance. Now, if you like, you can decide to *ignore* the components of lesser significance. You do lose some information, but if the eigenvalues are small, you don't lose much. If you leave out some components, the final data set will have less dimensions than the original. To be precise, if you originally have η dimensions in your data, and so you calculate η eigenvectors and eigenvalues, and then you choose only the first { eigenvectors, then the final data set has only { dimensions. What needs to be done now is you need to form a *feature vector*, which is just a fancy name for a matrix of vectors. This is constructed by taking the eigenvectors that you want to keep from the list of eigenvectors, and forming a matrix with these eigenvectors in the columns.

|

Feature Vector $=\{eig_1,eig_2,eig_3.....eig_n\}$

Step 5: Deriving the new data set

This final step in PCA, and is also the easiest. Once we have chosen the components (eigenvectors) that we wish to keep in our data and formed a feature vector, we simply take the transpose of the vector and multiply it on the left of the original data set, transposed.

Final Data=Row Feature Vector x Row Data Adjust

Where *Row Feature Vector* is the matrix with the eigenvectors in the columns *transposed* so that the eigenvectors are now in the rows, with the most significant eigenvector at the top, and Row *Data Adjust* is the mean-adjusted data *transposed*, i.e. the data items are in each column, with each row holding a separate dimension.

3. Feature Extraction Using LDA:

Algorithm:

Step1. Suppose there are R classes. Let $\boldsymbol{\mu}_r$ be the mean feature vector for class r. Let K_r be the number of training samples from class r. Let $K = \Sigma\, K_r$ be the total number of samples.

Step2. Calculate Within-class scatter matrix:

$$S_w = \sum_{r=1}^{R} \sum_{k=1}^{K_r} (x_k - \mu_r)^T (x_k - \mu_r)$$

Step3. Calculate Between-class scatter matrix:

$$S_b = \sum_{r=1}^{R} (\mu_r - \mu)^T (\mu_r - \mu)$$ Where $\mu = \frac{1}{R} \sum_{r=1}^{R} \mu_r$

Step4. LDA computes a transformation $\boldsymbol{V}$ that maximizes the between-class scatter while minimizing the within-class scatter:.

$$\text{Maximize} \frac{\det(V^T S_b V)}{\det(V^T S_w V)}$$

Such a transformation retains class separability while reducing the variation due to sources other than identity (e.g., illumination).

4. Classification Using Neural Network:

What is a Neural Network?

Neural networks consist of basic units modeled on biological neurons. Each unit has inputs that it combines into a single output. Neural networks are trained using examples and then they are run on data sets. Based on the training they receive they carry out activities.

Most widely used network is feed forward network with a back propagation learning algorithm.

Back Propagation Learning:

In 1969 a method for learning in multi-layer network, Back propagation (or generalized delta rule), was invented by Bryson and Ho**.]** Training the network is setting the best weights on the inputs of each units.

It consists of the following steps:

Network gets training example and using the existing weights in the network it calculates the output or outputs of the example. Back propagation calculates error by taking the calculated results and actual result. Error is fed back through the network and the weights are adjusted t minimize error.

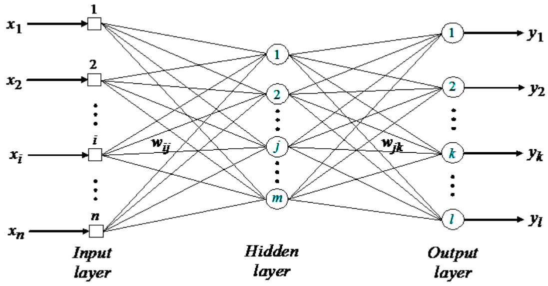

FIGURE :Neural Networks

We used a three layer perceptron neural network .i.e. input layer, hidden layer and the output layer, for classification of the input data. A simple back propagation algorithm is used to update weights according to desired values. Three layers MLP neural network will train using training face images and at its output layer, it produces a 10×1 vector that each elements of that vector is a number between zero and one representing similarity of input face images to each of ten classes. Training face image enter the neural network and according to their class, a back propagation error, spread on the network and correct the weights toward the right values. The input face image will classified to the class which has the greatest similarity to it. For example if for a test input face image, row 3 of network's output be greater than other rows, that test face images will classified to class 3.

Description of proposed method:

In this system we have large database i.e. training set. First we find average value by using image processing then we obtain average image & find difference images ϕ i. From that image find covariance matrix. By using covariance matrix, calculate Eigen vectors & generate eigen faces. Now sorting eigen value in such a way that greater information eigen value take first. Similarly compute eigen vector by using linear transformation of PCA & find projection of each classes i.e. U_i .This is projection of face space or basis of projection which is two dimensional.

Now we want to recognize test image A i.e. applying to that system. By image processing we find mean value of that image and multiply that mean image with each projections & generate that much feature which is in training set.

Now compare that features by using Ecludian distance. After comparing generate error value & select that features which give minimum value. By using threshold it decide that which features are selected with the help of error value & i.e. about input image.

Pre-processing:

In this part, at first we manually cut the input images to 40 X 40 images in order to remove the background information and have only face details (face images in most databases contain background information that is not useful for recognition, Figure (1) shows some face images with background information)

Fig. 1 Sample Face images that used in our experiment

After that we histogram equalize all input images in order to spread energy of all pixels inside the image and then normalize them to equalize amount of energy related to each face image.

As a next step, we subtract mean images from face images to mean center all of them. Fig. 2 shows mean images of face images that used in our experiments.

Fig. 2 Mean images related to our selected faces. Finally all preprocessed face images change to a vector (1600 X 1 vectors) and go to the next step. For complexity and memory size reduction, mean image is computed.

Dimensionality Reduction:

As mentioned in the previous section, we cropped every input image to 40×40 image; as a result the input of this stage is a preprocessed 1600×1 vector. We used these vectors to estimate the covariance matrix. After estimation of the covariance matrix, significant eigenvectors of the covariance matrix are computer. Number of eigen-vector depend to our application and accuracy that we need, it is clear that if we compute large number of eigen-vectors accuracy of the method improved but computational complexity increased in this step and next step. In this stage, we computed 100 most significant eigenvectors and related eigen-faces. By projection of every input image on these eigenafeces, they will convert to reduce size 100 × 1 vectors which will be go to LDA feature extraction part. We repeated our experiment with different values of significant eigen-vectors and choose them equal to 20, 40, and 60 and 80 and compared the performance of the proposed face recognition method.

LDA Feature Extraction:

Outputs of dimension reduction part are 100×1 vectors which are used to construct within class scatter matrix and covariance matrix. As mentioned in section II, significant eigenvectors of $\mathbf{S_w^{-1} S_B}$ can used for separability of classes in addition to dimension reduction. Using 100×1 vectors, $\mathbf{S_w^{-1} S_B}$ computed and then eigenvectors related to the greater eigenvalues are selected. In our experiment we considered 10 classes, therefore there are 9 major eigenvectors (Fisher faces) associated with non-zero eigenvalues which have separability capability. It is clear that extracting all of 9 LDA features increase the discriminatory power of the method. Therefore, this section produce 9×1 vectors which are used as input of MLP three layer neural network. We demonstrate operation of this part, at first covariance and within class scatter matrices are estimated and then significant eigenvector of $\mathbf{S_w^{-1} S_B}$ are computed (Fisher Faces).

IV. EXPERIMENTAL RESULTS

We applied the proposed new face recognition method on YALE face datasets for separation of ten classes. We cropped the input images to reduce their size to 40×40.

Fig.3 Pre-processed Sample face Images

Our selected database contains greyscale images of 10 subjects in GIF format. In these experiments, we considered 60 images per each subject (total 600 images) containing different illumination and different poses, which 40 images of each used for training and remaining 20 images used for

testing the method. Fig. 3 shows some of selected pre-processed subjects in different position and illumination. Then 100 significant Eigen faces are computed in stage 2, where Fig.3 shows first 50of them.

Table I
Comparison of Recognition Rate

	DCT-LDA-Short Dist 2004	DCT-2004	PCA-LDA 1994	PCA-Short Dist	Proposed Method
Recognition Rate	97.7	84.5	94.8	84	98

CONCLUDING REMARKS

In this paper, a new Face recognition method is presented. The new method was considered as a combination of PCA, LDA and neural networks. We used these algorithms to construct efficient face recognition method with a high recognition rate. Proposed method consists of four parts: i) image pre-processing that includes histogram equalization, normalization and mean centering, ii) dimension reduction using PCA that main features that are important for representing face images are extracted iii) feature extraction using LDA that significant features for class separability are selected and that classify input face images into one of available classes. Simulation results using YALE face datasets demonstrated the ability of the proposed method for optimal feature extraction and efficient face classification. In our simulations, we chose 10 persons and considered 40 training image and 20 test image for each person (totally 400 training and 200 test face images).

Experimental results show a high recognition rate equal to98.5% (in average one misclassification for each 200 face images) which demonstrated an improvement in comparison with previous methods. The new face recognition algorithm can be used in many applications such as security methods.

Reference:

[1] J. R. Solar, P. Navarreto, " Eigen space-based face recognition, IEEE Tran. , Systems man And Cybernetics- part c: Applications, Vol. 35, No. 3, 2005.

[2] O.Deniz, M. Castrill_on, M. Hern_andez, "Face recognition using independent component analysis and support vector machines" , Pattern Recognition letters, Vol. 24, PP. 2153-2157, 2003.

[3] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition", IEEE Trans. pattern Anal. Machine Intel., Vol. 24, No. 6, PP. 780-788, 2002.

[4] H. Othman, T. Aboulnasr, " A separable low complexity 2D HMM with application to face recognition" IEEE Trans. Pattern. Anal. Machie Inell., Vol. 25, No. 10, PP. 1229-1238, 2003.

[5] M. Er, S. Wu, J. Lu, L.H.Toh, "face recognition with radial basis function(RBF) neural networks", IEEE Trans. Neural Networks, Vol. 13, No. 3, pp. 697-710.

[6] K. Lee, Y. Chung, H. Byun, "SVM based face verification with feature set of small size", electronic letters, Vol. 38, No. 15, PP. 787-789, 2002.

[7] A. M. Martinez, A. C. Kak, "PCA versus LDA", IEEE Trans. Pattern Anal. Machine Intell, Vol. 23, pp. 228-233. 2004.

[9] S. Pang, S. Ozawa, N. Kasabov," Incremental linear discriminant analysis for classification of data streams", IEEE Trans. on Systems, Man and Cybernetics, vol. 35, no. 5, pp. 905-914, 2005.

[10] M.J.Er, W.Chen, S.Wu, "High speed face recognition based on discrete cosine transform and RBF neural network", IEEE Trans. On Neural Network, Vol. 16, No. 3, PP. 679,691, 2005.

[18] X. Y. Jing, D. Zhang, "A face and palm print recognition approach based on discriminant DCT feature extraction", IEEE trans. on Sys. Man & Cyb., Vol. 34, No. 6, PP. 2405-2415, 2004."