Comparison of Principal Component Analysis and Linear Discriminant Analysis for Face Recognition (March 2007)

P.E. Robinson, *University of Johannesburg, W.A. Clarke MIEEE, SAIEE*

***Abstract*—In this paper two Face Recognition techniques, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), are considered and implemented using a Nearest Neighbor classifier. The performance of the two techniques is then compared in facial recognition and detection tasks. The comparisons are done using a facial recognition database captured for the project that contains images captured over a range of poses, lighting conditions and occlusions.**

***Index Terms*—Face recognition, Eigenfaces, Fisherfaces, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA).**

I. Introduction

THE field of Face Recognition has received a lot of attention in recent history and as a result a very large variety of Face Recognition techniques have been documented in the literature. One of the main concerns that are currently faced is the objective comparison and benchmarking of the capabilities of these techniques and a number of empirical evaluation techniques and protocols have been put forward for this purpose such as the FERET evaluation protocol [1, 2].

The existing protocols use statistical methods and large testing databases that contain a large sample of facial images that spans the capture conditions that a facial recognition system is likely to face in a real-world application [1]. The evaluation methodology used in this paper makes use of statistical comparison and by selecting Face Recognition techniques that use identical classifiers a direct comparison of recognition performance under a variety of image capture conditions which are contained in the facial image database captured specifically for this project.

One of the key components of Facial Recognition techniques is the method of facial feature extraction used. The two main approaches to facial feature extraction are a holistic template matching approach and the extraction of local geometrical features [3]. The techniques used in this paper are based on holistic template matching.

Manuscript received March 2, 2007.

P.E. Robinson is with the University of Johannesburg, Kingsway, Auckland Park, Johannesburg, South Africa (e-mail: thedescant@gmail.com).

II. Facial Recognition Techniques

A. Principal Component Analysis (PCA)

Principal component analysis is a well known method used to approximate a set of data with lower dimensional feature vectors. In the case of Face Recognition the data considered is an 8-bit grayscale image which is converted into a vector in a column-wise fashion [4].

The first stage of the PCA system is the training stage. A set of facial images which is made up of classes of images of subjects that should be recognized by the system is used as a training set. The training set is used to create a covariance matrix of the training data whose strongest eigenvalues will form the basis of the vector space spanned by all the training faces which is called the Face Space. If each face image is defined as Γ_1, Γ_2, Γ_3 etc then the average face is defined as $\Psi = \frac{1}{M}\sum \Gamma_i$ where M is the number of face images and Γ_i has a length of N [4].

Each face in the ensemble differs from the average by $\Phi_i = \Gamma_i - \Psi$ and then we can then find the covariance matrix C as follows [4]:

$$C = \frac{1}{M}\sum \Phi_n \Phi_n^T = AA^T \text{ where } A = [\Phi_1 \Phi_2 \ldots] \quad (1)$$

This means C is an N^2xN^2 matrix and finding the eigenvectors of this is a time consuming task. In [4] a trick is used to avoid this task but it was found that the use of Singular Value Decomposition (SVD) was computationally faster. SVD is then used to find the M' strongest eigenvectors of C, defined as matrix *u,* which due to their face like appearance are called the Eigenfaces [4].

The Eigenfaces found in this way are the basis for the Face Space and all the face images from the training set are projected into the face space in the following manner:

$$\omega_k = u_k^T(\Gamma - \Psi) \text{ where k = 1,…,M'} \quad (2)$$

The resultant weight vector ω represents the face's position in the Face Space [4].

The second stage of the PCA system is the recognition stage. This stage is used to recognize unknown faces. The probe face which is to be recognized is converted into a vector in a column-wise fashion and is projected into the Face Space using operation (2). The result is the position of the probe image in the Face Space and now a Nearest Neighbor classifier is used (merely the Euclidian distance between the probe image and other faces in the Face Space) to find if the probe face is close enough to the training images in the face space to be considered a known face and then which face class is the closest to the probe image. The closest face class is then considered the identity of the unknown face [4].

B. Linear Discriminant Analysis of Principal Components (PCA + LDA)

The LDA + PCA algorithm is also a dimension reduction algorithm like the pure PCA algorithm. The general procedure behind the algorithm is exactly the same as that described above for the PCA algorithm. The primary difference between the algorithms is that the eigenvector analysis is applied to a Separation Matrix derived from Fisher's Linear Discriminant function instead of the Covariance matrix. One of the main reasons for the selection of this algorithm is the fact that the similarity measures (Nearest Neighbour Classifier) are directly comparable allowing for easier comparison of performance results [5].

LDA does not just use the training set plainly to find the basis of the Face Space. LDA takes into account the relationships between the various training images and their relationship to the training set as a whole. The first stage in the LDA algorithm is to label the training set data, specifying that all images of a single subject are in the same Class and the images of every subject in the training set are in different Classes [5].

Once the training set is thus labeled we are in a position to use Separation Matrix to perform a cluster separation analysis. The goal is to minimize Within-Class variance (keep images of the same individual closely clustered) and maximize Between-Class separation so it becomes easier to distinguish between the various class clusters [5, 6].

The Separation Matrix we are going to use for the eigenvector analysis is given by $S = S_w^{-1} S_b$ where S_w is the within-class scatter matrix and S_b is the between class scatter matrix.
The scatter matrices are given by:

$$S_w = \sum_{i=1}^{L} \Pr(C_i) \Sigma_i \tag{4}$$

and

$$S_b = \sum_{i=1}^{L} \Pr(C_i)(\mu - \mu_i)(\mu - \mu_i)^T \tag{5}$$

Where $\Pr(C_i)$ is the probability of Class i occurring, μ is the Sample Mean, μ_i is the Class mean for Class i and the Average Scatter $\Sigma_i = (V - \mu_i)(V - \mu_i)^T$ where V are the image vectors of Class i [5, 6].

The primary reasons that PCA is used in conjunction with LDA are as follows:

- Pure LDA algorithms have documented problems with detecting face images that were not in the training set.
- The $S = S_w^{-1} S_b$ operation can be very expensive in terms of processing time if the scatter matrices are too large due to the matrix inversion in the calculation.

Using the PCA projections helps solve both these problems. By working in the PCA subspace, unknown faces are projected into the general area of the training faces in the *Face Space* allowing the LDA to more easily classify them. The projection of the scatter matrices into the PCA subspace also reduces their dimension, which improves the time taken to perform the calculation to find S [5, 6].

Now to include PCA in the algorithm we first find the Eigenfaces for the training set before beginning the LDA calculations. Once the two scatter matrices have been calculated using (4) and (5) they are projected into the PCA subspace by the following operation [5]:
$S_{w(PCA)} = u^T S_w u$ and $S_{b(PCA)} = u^T S_b u$ where u are the Eigenfaces.

Then the Separation matrix is calculated using $S = S_{w(pca)}^{-1} S_{b(pca)}$ and the Eigenvalues and Eigenvectors of S are found. The top Eigenvectors E are then projected back into full space as follows: $E_{(full\ space)} = \underline{u}.E_{(PCA)}$. The resulting Eigenvectors form the basis of the *Face Space* and are called Fisherfaces and are used in exactly the same procedure for recognition tasks as the Eigenfaces [5].

III. Experiments

A. Test Methodology

The test procedure was designed to compare the performance of both algorithms when faced with variations in illumination and pose and their ability to detect the presence of a face in an image.

1) Image Datasets

To perform the desired tests a number of datasets had to be compiled. These data sets had to contain facial images captured in a variety of illumination condition and pose angles.

The primary data set (UJ Face Database) was captured specifically for the project and contained a variety of subjects from different ethnicities to try and create a diverse sample of the population.

The dataset contained the following images of each subject:

- Average uniform lighting

- Average uniform lighting with sunglasses
- Average uniform lighting with sunglasses and a hat
- Dim uniform lighting with no occlusions
- Bright uniform lighting with no occlusions
- Average lighting from left
- Average lighting from right.

This set of images was captured at three pose angles of 0°, 20° and 40° for each subject. The primary dataset contained 651 facial images.

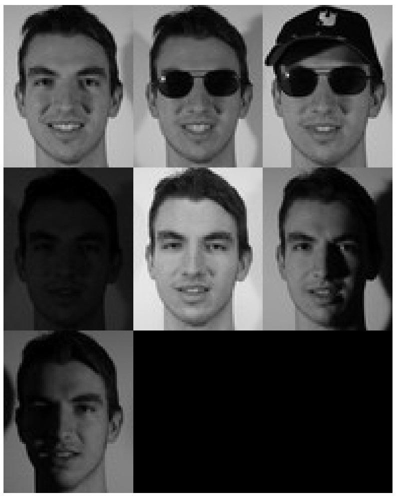

Fig. 1. Typical image set from 0°

For the pose variation tests the UMIST database was used which contained 22 images per subject ranging from frontal to profile views [7].

The final dataset was required for the detection tests and comprised of 20 known faces, 20 unknown faces, 20 animal faces, 20 small objects and 20 large object.

2) Preprocessing

The purpose of the Preprocessing is to format the images in the test image sets to be suitable for the algorithms to use. This is done in the following steps:

- The images are converted to 8-bit grayscale intensity images.
- The images are cropped to contain only the subjects head as shown in Fig 1.
- The images are resized to 64x80 pixels.

3) Pose variation test methodology

The purpose of this set of tests was to investigate the two algorithms sensitivity to pose variations. For this test the UMIST face database [7] was used and from the database 16 subjects sampled over 22 different pose angles, from almost profile to full frontal view, were selected.

From each subject 11 photos were chosen as probes and 11 for training. The training and probe sets where interleaved so they both included a range of poses from full frontal views through to almost profile views. The experiment then proceeded to use varied training set sizes (from 2 up to 11) to train the algorithms and probe them using all 11 probe set images.

4) Illumination variation test methodology

The purpose of this test set was to investigate the effects of Illumination variations on performance of the two algorithms and whether pose variations affect the algorithms ability to deal with illumination variations. The UJ Face Database was used in this experiment and the 9 average lighting images per subject were used as a training set and all the illumination variations over the three pose angles were used as the probe set.

The training set was increased from 2 images through to 9 images during the course of the test set and for each training set size the algorithms were probed with the entire probe set.

5) Detection test methodology

The purpose this set of tests was to determine the effectiveness of the two algorithms for the purpose of face detection and how training set size can influence their effectiveness.

In both algorithms we can use the distance from the Face Space of the probe image as a measure of its likeness to a face. We have three classes into which an image could fall. Firstly an image could be a known face meaning it is an image of one of the subjects used in the training of the algorithm. Secondly the image could be of an unknown face which is a human face that is not an image of one of the subjects in the training set. Finally an image could be far enough away from the Face Space so as not to be a human face at all. These three areas are separated by distance thresholds.

The first test in the set will be used to determine average distance between the three classes of probe image and the training set. This will allow us to find the thresholds separating the 3 classes for each algorithm. This is done by training the algorithms with a set of faces from the UJ face database and then probing the algorithms with 20 images from each of the 5 image sets in the detection image set (known faces, unknown faces, animal faces, small object and large objects).

The second test uses these thresholds to test the accuracy of the algorithms at detecting which of the 3 classes a probe image falls within with the size of the training image set ranging from 1 to 20 images.

B. Test Results

1) Pose variation test

In this test a series of recognition rate results for each pose angle from frontal to profile was obtained for training sets containing 2 images through to 11 images. Fig 2 shows the results of the PCA and LDA + PCA algorithms for a training set size of 6 images. From this figure it is apparent that the LDA + PCA algorithm performs better than the PCA algorithm over a broader range of pose angles.

However when the training set size increased to 11 images the results of the two algorithms converged and became very similar. This tells us that the LDA + PCA algorithm performs better than the PCA algorithm when dealing with extreme pose angles when it's training set is sparse.

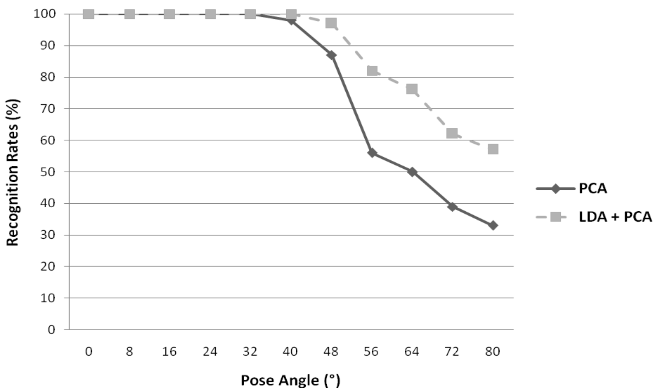

Fig. 2. Recognition rates of both algorithms for pose angles from frontal tc profile. These results are for a training set size of 6 images.

The results from the Illumination variation test as can be seen in Fig 3 which shows the results for each of the 4 lighting conditions in the UJ Face database for each of the 3 pose angles. It can be seen that both algorithms are extremely sensitive to illumination variations. Uniform bright illumination does not pose much of a problem for the algorithms but when shadows obscure the face like in the dim, left lit and right lit cases the algorithms do not perform well. However it can be seen that the LDA + PCA algorithm out performs the PCA algorithm when faced with illumination variations.

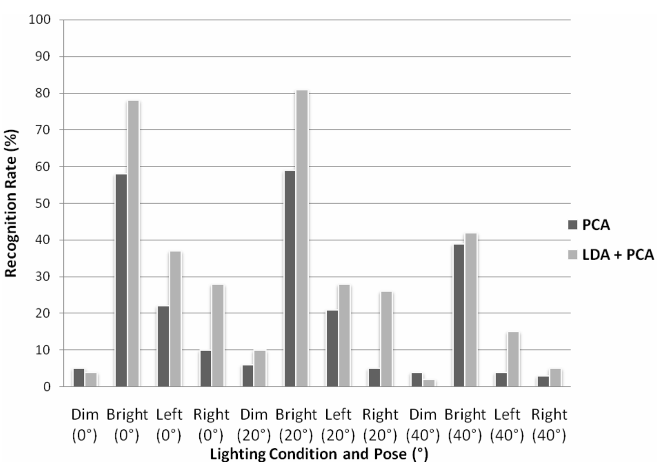

Fig. 3. Recognition rates of both algorithms for the labeled illumination conditions and pose angles. These results are for a training set size of 6 images.

3) Detection test

The first part of the detection test used a fixed training set for each algorithm and then probed the algorithms with various classes of image. From the average minimum distances of the various classes to the face space the thresholds defining the zones in the face space that contains the known faces (center zone), unknown faces (second zone) and non human faces (outer zone) where found and are shown in Fig 4. As can be seen in the figures the LDA + PCA zones are far more compact that the pure PCA zones. This is to be expected as the point of the Discriminant function in the LDA part of the algorithm is to minimize scatter of images that are human faces and faces in the known class which can be seen [6]. This means that there is less room for variance when performing detection tasks.

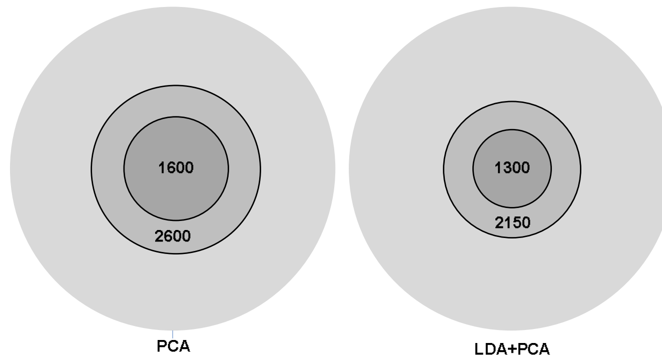

Fig. 4. The thresholds found for the various classes in the detection tests. The classes are represented as follows: known faces (center zone), unknown faces (second zone) and non human faces (outer zone).

As can be seen in Fig 5 LDA + PCA does not perform as well as just PCA for detection of images that fall into the various zones. This is due to the smaller tolerance for variance that the smaller thresholds cause in the LDA portion of the algorithm. This intolerance results in inaccuracy when performing these detection results especially when the tr65
aining set sizes are sparser and the LDA Space is more tightly clustered due to the Discriminant function.

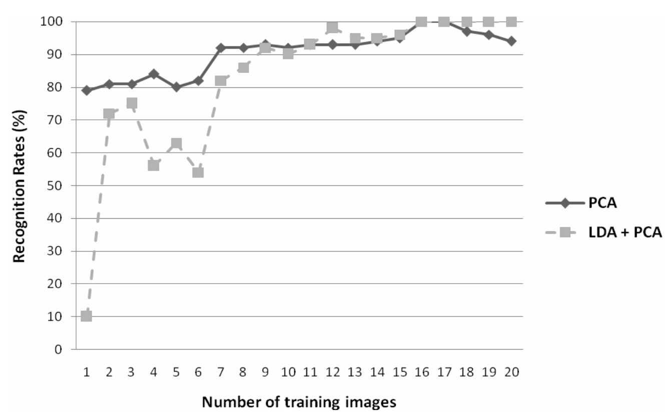

Fig. 5. The recognition rates of images that are in the Known Face class for a range of training set sizes

The effects of over training can also be noted in the PCA recognition rates from Fig 5 where there is a dip in the recognition rates for the 3 largest training set sizes. This is due to the system becoming over trained and the known face class zone becomes more compact and becomes intolerant to the variance in the detection operation.

IV. CONCLUSION

In conclusion the PCA and LDA + PCA algorithms were implemented and their performance was tested when faced with images of varying pose angle and illumination conditions. It was found that the LDA + PCA algorithm performed better than the PCA algorithm in recognition tasks especially when the training set was sparse. Both algorithms were very sensitive to illumination variations but both could cope well with a variety of pose angles as long as the algorithms were trained appropriately. Despite the LDA + PCA algorithm's better recognition performance its processing time is far greater than the PCA algorithm's which will influence the choice to use it over the PCA algorithm.

The detection tests were done to find the thresholds for the zones of the various classes of image. The results showed that for detection tasks the compact detection zone of the known and unknown face classes in the LDA + PCA algorithm made the algorithm intolerant to the variance of the detection operation. Due to this the PCA algorithm performed better at face detection tasks than the LDA + PCA algorithm.

REFERENCES

[1] W. Zhao, R. Chellappa, A. Rosenfeld, PJ Phillips, "Face Recognition: A Literature Survey", *ACM Computing Surveys*, 2003, pp. 399-458.

[2] P.J. Phillips, P.J. Rauss, SZ Der, "The FERET evaluation methodology for face recognition algorithms", IEEE Transactions on Patterns Analysis and Machine Intelligence, Vol. 22, No. 10. 2000.

[3] C. Podilchuk, A. Patel, A. Harthattu et al, "A New Face Recognition Algorithm using Bijective Mappings", *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, 2005, pp. 165.

[4] M Turk, A Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neurosicence*, Vol. 3, No. 1, 1991, pp. 71-86

[5] W. Zhao, R. Chellappa, A. Krishnaswamy, "Discriminant Analysis of Principal Components for Face Recognition", *Proc. of the 3rd IEEE International Conference on Face and Gesture Recognition*, FG'98, 14-16 April 1998, Nara, Japan, p. 336

[6] K. Etemad, R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images", *Journal of the Optical Society of America A*, Vol. 14, No. 8, August 1997, pp. 1724-1733

[7] D. Graham, N. Allinson. "Face Recognition: From Theory to Applications", NATO ASI Series F, Computer and Systems Sciences, Vol. 163. pp 446-456, 1998.

Copyright Information