

# Comparison of PCA and MPCA with Different Databases for Face Recognition

Ambika.D

School of Inf. Technology  
&Engg  
VIT University  
Vellore, India

Arathy.B

School of Inf. Technology  
&Engg  
VIT University  
Vellore, India

Srinivasa Perumal.R

School of Inf. Technology  
&Engg  
VIT University  
Vellore, India

## ABSTRACT

Face recognition is one of the Biometric characteristics for person identification. In this paper, Face recognition is done using two feature extraction techniques PCA (Principal Component Analysis) and MPCA (Modular Principal Component Analysis). PCA is a linear projection method in which dimensionality reduction is applied to the original image space. MPCA is an improved version of PCA in which each image (Face image) is divided into number of sub-block image and then PCA is applied for each sub-block image. The experimental result shows the accuracy of PCA and MPCA for different database images.

## General Terms

Pattern recognition; Biometrics.

## Keywords

PCA; Modular PCA; Face Recognition; Eigen Faces.

## 1. INTRODUCTION

Biometrics is the automatic recognition of humans using their physical or behavioral characteristics. In Biometrics, it is easy to verify and validate the identity of the individual. Various biometric techniques like Face, DNA, Fingerprint, Gait, Iris, Palm print, Voice are implemented recently.

Face Recognition is a unimodal biometric system for image identification and verification. It is carried out using many techniques like Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA), Independent Component Analysis (ICA), Modular Principal Component Analysis (MPCA) and 2-Dimensional Principal Component Analysis (2DPCA). PCA overcomes the high dimensional pattern recognition problem using decomposition technique to reduce dimensionality. In MPCA it is easy to recognize the face image even if the image has varying facial expressions, pose variations, illumination changes, etc.

Chellappa made a survey for face recognition based on features and neural networks [1]. PCA is done for frontal face images using best Eigen vectors called Eigen faces [2][3]. In Eigen face approach, after dimensionality reduction, the Euclidean distance is measured for the face images [4]. In MPCA, the accuracy is higher for face images having different lighting variation and changes in facial expression [5]. In improved MPCA, weighted average is applied for each and every sub-block and then PCA technique is followed [6]. PCA technique is applied to get higher accuracy rate for face recognition and the performance is evaluated. KPCA is also carried out in [7]. Face Recognition in subspaces and feature extraction for each subspace is followed in [8]. A survey of many face recognition techniques is followed in [9].

The objective of this paper is to enhance the recognition rate for varying facial expression and illumination changes. PCA is a simple and efficient method to perform face recognition. But the recognition and accuracy rate is not high for images having different expressions and poses. In this paper, accuracy rate is found for MPCA and it is compared with PCA.

## 2. PCA

PCA is the most simple and efficient method for face recognition. It is a decomposition technique to reduce dimensionality (2 Dimensions). It is mainly used for image compression which is also known as Karhunen-Loeve Transform. PCA overcomes the high dimension pattern recognition problem in Face recognition by using Eigen value decomposition technique. In PCA, Holistic approach is used for feature extraction to extract features globally.

### 2.1 Steps for EigenFace Algorithm

#### 2.1.1 Data Acquisition

Face images are captured and stored in database. The collected images are pre-processed using histogram equalization and they are used for further processing. All the images are concatenated to form a single matrix ( $N^2 \times M$ ) where M represents the number of images.

#### 2.1.2 Subtract the Mean

Initially mean value is calculated for the input matrix and this value is subtracted from the original matrix ( $\Gamma_i$ ). From this the mean adjusted value is calculated. The mean ( $\Psi_i$ ) and mean adjusted value ( $\Phi_i$ ) is reshaped to form the average image and mean adjusted image.

$$\psi = 1/M \sum_{n=1}^M \Gamma_n.$$

$$\Phi_i = \Gamma_i - \psi. \quad (1)$$

#### 2.1.3 Covariance Matrix

In this step, the covariance matrix C is calculated as,

$$C = 1/M \sum_{n=1}^M \Phi_n \Phi_n^T. \quad (2)$$

#### 2.1.4 Eigen values and Eigen vectors

From covariance matrix, Eigen values and Eigen vectors are found using Jacobian method. The outcome of covariance is

symmetric, hence Eigen values & vectors always are symmetric. Eigen faces are calculated as,

$$U_k = \sum_{k=1}^M V_k \Phi_k. \quad (3)$$

### 2.1.5 Recognizing the Face

In this step, weights ( $W_k$ ) should be calculated for both train and test set. The Euclidean distance is found by calculating the difference in weights. A threshold value should be set for recognition. If the Euclidean distance is equal to the threshold value, then it is known image having same expression. If it is less than the threshold value, then it is known image having different expression. If it is greater than the threshold value, then it is unknown image which is not recognized image. The weight vector ( $\Omega^T$ ) is calculated as,

$$W_k = U_k(\Gamma_i - \Psi_i).$$

$$\Omega^T = [W_1 W_2 W_3 \dots W_M]. \quad (4)$$

The normalized training set is represented as,

$$\Omega_i = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_k \end{bmatrix}. \quad (5)$$

## 3. MODULAR PCA

In PCA, recognition rate is higher only for frontal face images. The information of the face is taken globally for processing and checked whether it is recognized or not. Though PCA overcomes the high dimensionality pattern recognition problem, the accuracy rate is not high. In case of images having different poses, facial expressions and change in illumination, the recognition rate of PCA will not be high.

In MPCA, initially the face images are partitioned into N number of sub-block images and then apply steps of PCA for each and every sub-block image by using local information of the face. Here the first sub-image of test image is taken initially and compared with all the images in the database. The images matched with this first sub-image are selected. Again, the second sub-image is checked with this set of selected images and then the matched images are found. Further, the third sub-image is compared with the previous set of selected images and the same procedure is applied to find the recognized image. Finally the last sub-image is also compared and the image is recognized. If it got rejected in the first step, then the image is considered as not recognized. MPCA overcomes the problems of PCA under the varying conditions of pose variations, facial expressions, and illumination changes. So we expect that the accuracy rate will be better for Modular PCA.

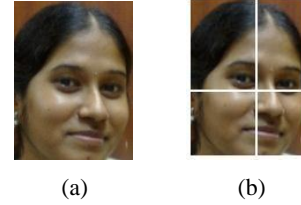
### 3.1 Steps for Modular PCA

#### 3.1.1 Divide into sub-blocks

The training set images are divided into N number of sub-blocks. L represents the size of the image. The size of each sub-blocks are divided as  $L^2/N$ . The sub-blocks can be found as,

$$I_{ij}(m,n) = I(L/\sqrt{N}(j-1) + m, L/\sqrt{N}(j-1) + n) \forall i, j. \quad (6)$$

Where i vary from 1 to M, M is the number of images in the training set, j varies for 1 to N, N represents number of sub-images, m and n varies 1 to  $L/\sqrt{N}$ . Here, each image is divided into four sub images (N=4) is shown in Fig. 1.



**Figure 1 a) original image b) image divided into N sub-blocks (N=4)**

#### 3.1.2 Data Acquisition

In this step, sub-images ( $I_i$ ) are placed in training set. The collected images are pre-processed using histogram equalization and they are used for further processing. All the images are concatenated to form a single matrix ( $N^2 \times M$ ) where M represents the number of images.

#### 3.1.3 Average Image

The average image A of the sub-images are found as,

$$A = 1/M \cdot 1/N \sum_{i=1}^M \sum_{j=1}^N I_{ij}. \quad (7)$$

#### 3.1.4 Normalize the image

Normalization can be done by subtracting the matrix with its mean.

$$Y = I_{ij} - A \forall i, j. \quad (8)$$

#### 3.1.5 Covariance matrix

The covariance matrix is found from the normalized sub-images as follows,

$$C = 1/M \cdot 1/N \sum_{i=1}^M \sum_{j=1}^N Y_{ij} Y_{ij}^T. \quad (9)$$

#### 3.1.6 Eigen values and Eigen vectors

From covariance matrix, Eigen values and Eigen vectors are found using Jacobian method. The outcome of covariance is symmetric, hence Eigen values & vectors always are symmetric.

$$U_k = \sum_{k=1}^M V_k \Phi_k. \quad (10)$$

Where  $U_k$  represents Eigen faces,  $V_k$  represents Eigen vectors and  $\Phi_k$  is mean adjusted value.

### 3.1.7 Recognizing faces

In this step, weights ( $W_k$ ) should be calculated for both train and test set. The Euclidean distance is found by calculating the difference in weights. A threshold value should be set for recognition. If the Euclidean distance is equal to the threshold value, then it is known image having same expression. If it is less than the threshold value, then it is known image having different expression. If it is greater than the threshold value, then it is unknown image which is not recognized image. The weight vector ( $\Omega^T$ ) is calculated as

$$W_k = U_k(\Gamma_i - \Psi_i).$$

$$\Omega^T = [W_1 W_2 W_3 \dots W_M]. \quad (11)$$

The normalized training set is represented as,

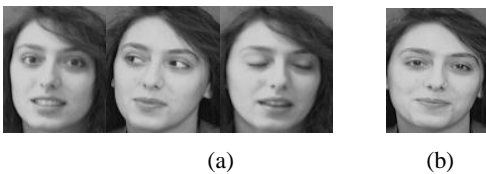
$$\Omega_i = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_k \end{bmatrix}. \quad (12)$$

## 4. FACE DATABASES

For both PCA and MPCA techniques, we have used three databases- ORL database, JAFFE database, and Indian database.

### 4.1 ORL Database

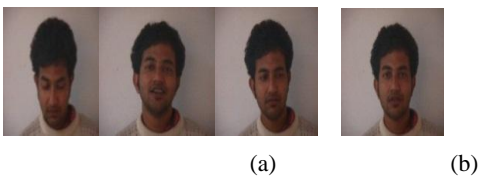
In ORL database, forty distinct subjects are taken in the database and ten images having different expressions of each subject are considered. Fig. 2(a) shows the sample of test images and Fig. 2(b) shows the sample of train images in ORL Database.



**Figure 2** (a) training image and (b) test image in ORL database.

### 4.2 Indian Database

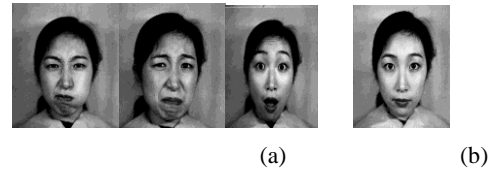
In Indian database, forty distinct subjects are taken in the database and each image having different expressions of each subject are considered. Fig. 3(a) shows the sample of test images and Fig. 3(b) shows the sample of train images in Indian Database.



**Figure 3** (a) training image and (b) test image in Indian database.

### 4.3 JAFFE Database

In JAFFE database, ten distinct subjects are taken in the database and ten images having different expressions of each subject are considered. Fig. 4(a) shows the sample of test images and Fig. 4(b) shows the sample of train images in JAFFE Database.



**Figure 4** (a) training image and (b) test image in JAFFE database.

## 5. EXPERIMENTAL RESULTS

The performance of both PCA and MPCA techniques are tested by varying the number of eigenvectors. The Euclidean distance is calculated by finding the difference between the weights of test and train sets. A threshold value has been set to increase the recognition rate.

For PCA, the Euclidean distance is calculated for both test and train set in the database. If the Euclidean distance is equal to the threshold value which we have set, then the input image completely matches with the processed image and is found to be recognized image having same expression is shown in Fig. 5.



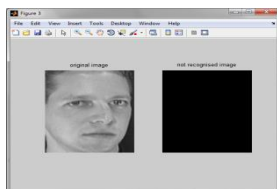
**Figure 5** Recognized image of same expression in Indian Database for PCA

Another test was done for an image which was present in the database but having different face expression. If the Euclidean distance is less than threshold value which we have set, then the input image matches with the processed image and is found to be recognized image having different expression is shown in Fig. 6.



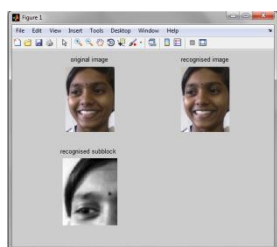
**Figure 6** Recognized image of different expression in Japanese Database for PCA

If the Euclidean distance is greater than threshold value which we have set, then the input image does not match with the processed image and hence it is an unknown face and is shown in Fig. 7.



**Figure 7** Image not recognized using ORL Database for PCA

For MPCA, the Euclidean distance of test image from each sub-image in the database is calculated. If the Euclidean distance is equal to the threshold value which we have set, then the sub-image completely matches with its original image and is found to be recognized image having same expression is shown in Fig. 8.



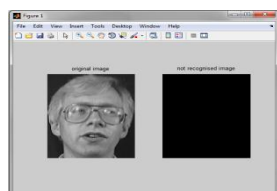
**Figure 8** Recognized image of same expression of Indian Database for MPCA

Another test was done for an image which was present in the database but having different face expression. If the Euclidean distance is less than threshold value which we have set, then the input image matches with the processed image based on the sub-image and is found to be recognized image having different expression is shown in Fig. 9.

If the Euclidean distance is greater than threshold value which we have set, then the input image does not match with the processed image and hence it is an unknown face and is shown in Fig. 10.



**Figure 9** Recognized image of different expression of Indian Database for MPCA



**Figure 10** Image not recognized in ORL Database for MPCA

In Table-1, the results of PCA and MPCA are calculated for three different databases- ORL Database, Indian Database, JAFFE Database. The metrics is calculated based on False Acceptance Rate (FAR), False Rejection Rate (FRR) and

Accuracy (ACC). FAR refers to the ratio of number of false acceptances divided by the number of attempts identified. FRR refers to the ratio of number of false rejections divided by the number of attempts identified. ACC is calculated based on FAR and FRR.

## 6. CONCLUSION

Face Recognition has several feature extraction method, by comparatively PCA and Modular PCA with three different datasets. By comparative study we understand modular PCA provides better result than the PCA base on accuracy rate. The experimental result shows how it is perform better more than PCA in three different dataset. In large variation of illumination and expression of the face, the modular PCA will be useful for identification of the face but the computational complexity is higher.

**Table .1**Experimental results of different databases

Metrics	ORL Database		Indian Database		JAFFE Database	
	PCA	MPCA	PCA	MPCA	PCA	MPCA
<b>FAR (%)</b>	0.50	0.40	0.40	0.30	0.40	0.30
<b>FRR (%)</b>	0.40	0.40	0.30	0.30	0.35	0.20
<b>ACC (%)</b>	99.10	99.20	99.3	99.40	99.2	99.50

- [1] REFERENCES Chellappa, R., Wilson, C.L., Sirohey, S., 1995. Human and machine recognition of faces: A survey. Proc. IEEE 83 (5), 705-740.
- [2] Kirby and L.Sirovich "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces", IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol-12, No-1, January 1990.
- [3] M.Turk and A.Pentland, "Face Recognition using "Eigen Faces", Proc.IEEE Conf. on Computer Vision and Pattern Recognition. Pages 586-591, 1991.
- [4] Sheifali Gupta et al."A Bespoke Approach for Face Recognition using PCA" International Journal on Computer Science and Engineering (IJCSSE) , Vol. 02, No. 02, 2010 155-158.
- [5] Rajkiran Gottumukkal, Vijayan K.Asari, "An improved face recognition technique based on modular PCA approach", Pattern Recognition Letters 25(2004) 429-436.
- [6] Chengmao Han, "Modular PCA Face Recognition based on Weighted Average ", CCSC journal, November 2009, Vol-3, No. 11.

- [7] Kiran Jain & Sukhvir Singh, “Performance Evaluation Of Face Recognition Using PCA”, *International Journal Of Information Technology and Knowledge Management*, July-December 2011, Volume 4, No. 2, pp. 427-430.
- [8] Stan Z.Li & Anil K.Jain, ”Handbook of Face Recognition”, second edition. Springer, 2011.
- [9] Rabia Jafri and Hamid R.Arabnia, “A Survey of Face Recognition Techniques”, *Journal Of Information Processing Systems*, Vol.5, No.2, June 2009.