

# **Illumination and Expression Invariant Automatic Human Face Recognition using Wavelet, Eigen and Fisher Analysis**

**Shubhankar De**  
School of Education Technology  
Jadavpur University  
Kolkata, India

**Ranjan Parekh, Ph.D**  
School of Education Technology  
Jadavpur University  
Kolkata, India

## **ABSTRACT**

This paper studies recognition of human faces using wavelet transform, Eigen space mapping and Linear Discriminant Analysis/Fisher Analysis (LDA). Histogram Equalization is chosen as a preprocessing step to reduce the effect of variation in illumination on human faces. The preprocessed faces are then subjected to second level wavelet (Haar) decomposition for further calculation. Feature extraction is performed using Eigen space mapping followed by LDA on the second level approximation matrix (LL sub band). Manhattan distance is used as a classifier. The proposed scheme is tested on illumination and expression variant different face databases for performance evaluation.

## **General Terms**

Human Face Recognition, Biometrics

## **Keywords**

Histogram Equalization, Wavelet Transform, Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA)/Fisher Analysis

## **1. INTRODUCTION**

Over the last few decades intensive research work has been performed in the field of biometrics to identify humans as accurately as possible based on biological traits like fingerprint recognition, iris recognition, retinal scanning, palm print recognition etc. Face recognition is one of the popular techniques. A face recognition system identifies a human individuals based on the frontal face image and has applications in defense, security, automatic access control, human computer interface etc. However face recognition is still a challenging task which needs to deal with varying illumination condition, poses, facial expression, partial occlusion and aging. Generally face recognition has three steps (1) preprocessing, (2) data modeling and (3) classification. In the preprocessing step includes segmentation of the face image eliminating the background.. The second step is divided into two groups namely feature based and holistic based [1]. Feature based methods rely on facial features such as eyes, nose, chin and mouth which are analyzed to establish the relationship and positions between them[2]. In holistic approaches the entire face image are analyzed as a whole. A mathematical model is created to represent the image in low dimension, so it can be compared with other images for processing. Finally the third step involves similarity measurement using an appropriate classifier. The organization of this paper is as follows, section 2 throws light on the related works on face recognition, followed by section 3 explaining the proposed technique, section 4 provides details of the datasets and experimental results obtained. Section 5 provides the analysis of current approach

vis-à-vis other contemporary works and section 6 provides the overall conclusions and scopes.

## **2. PREVIOUS WORK**

One of the earliest works on face recognition is recorded in [3]. The problem was to select from the database a small set of records such that one of the image records matched the photograph. In [4] Eigen space was used for face recognition. As mathematical algorithms utilizing Eigen vectors were used for representation of primary components of the face, it is termed as Eigen face technique. It creates a unique weight for each face image, which are used to represent the Eigen face features. Hence comparison of these weights permits identification of individual faces from database. In [5] linear discriminant analysis approach is used to maximize the scatter between different classes and minimize the scatter between input data in the same class. In [6] a comparison between Fisher face and Eigen face was presented which concluded that Fisher face can handle more lighting variations and expressions than Eigen face. Face recognition approach using 2D PCA in template matching is proposed in [7]. Modular PCA is applied in [8]. Here each face image is divided into sub images and PCA approach is applied to each of these sub images. ICA (Independent Component Analysis) based feature extraction method is used in [9]. In [10] an entropy based high informative zone is selected and local feature is calculated using wavelet transform. In [11] a technique is proposed to design a facial recognition system using multilayer perceptions and probabilistic neural network based classifier. A Discrete Cosine Transform based approach was proposed in [12] for face recognition. In [13] a wavelet LDA based algorithm was proposed for face recognition. In [14] Sobel face and 2DPCA was proposed to reduce the illumination effect. In [15] A Gabor LDA based face recognition technique was proposed. Gabor Wavelet was used as a preprocessing step and then LDA was applied for feature extraction with Artificial Neural Network (ANN) as a classifier. In [16] age invariant face recognition scheme was proposed using self principle component analysis considering only on the eye region of a face because it is most stable part of the face which changes very slowly with respect to time. In [17] fully automatic face recognition was performed using Transform- invariant PCA. In [18] A Facial Expression recognition system was introduced using PCA to classify human emotions. In [19] Gabor Wavelet, PCA and Local Binary Pattern (LBP) based facial expression recognition algorithm was proposed where PCA or LBP were used for

dimension reduction. In [20] a scalar based template was proposed for face recognition in reduced Eigen plane.

### 3. PROPOSED METHODOLOGY

This paper proposes a system for recognizing human faces by considering variations in illumination conditions and facial expressions. To handle the varying illumination condition we consider Histogram Equalization as a preprocessing step. On the preprocessed face Haar wavelet decomposition is performed and second level approximation coefficient matrix is chosen to Calculate Eigen space features using PCA, retaining as much as possible variation in dataset. The objective of LDA is to find out the optimal transformation matrix so the ratio of between class scatter matrix and within class scatter matrix reaches to its maximum and imposed on the previously calculated Eigen space feature to calculate LDA features. Classification is done using Manhattan distance.

#### 3.1 Preprocessing

In this part, at first we manually crop the input images to a standard size of 100×100 pixels images in order to remove the background information and isolate only face details. Then it is converted to gray scale image (for color face databases only). An image histogram is a graphical representation of the tonal distribution in a digital image. The lighting variation is handled by a histogram equalization (HE) in order to distribute energy of all pixels inside the image and then normalize them to equalize amount of energy related to each face image.

#### 3.2 Wavelet Transform

Wavelet Transform (WT) has been a very popular tool for Image analysis in the recent years. The advantages of WT, Such as good time and frequency localizations, have been utilized in various areas like image compression, image Processing, computer graphics etc. By decomposing an image using WT, the resolutions of the sub-band images are reduced.

An image is decomposed into four sub bands using a 2D WT as shown in Figure 1. The band LL is a coarser approximation to the original image. The bands LH and HL record the contours/edges along horizontal and vertical directions respectively. While the HH band records the diagonal edges of the image. This is the first level decomposition. Further decomposition can be conducted on the LL sub band.

LL	LH
HL	HH

Fig1: 2D DWT first level decomposition

After second level decomposition of all the face images, LL sub band is chosen for further processing. The reasons are, (1) It reduces the computational complexity, (2) It contains most of the variance/energy of the original image after applying the wavelet transform [13].

#### 3.3 Mapping to Eigen Space

The approximation matrix/LL sub band of size (R x C) pixels, where R is number of rows and C is number of columns. It is

then transformed into a single column vector,  $\Gamma$ , of size (Q x 1) where  $Q = (R \times C)$ .

In the Training phase training matrix is formed by appending column vector of each approximation matrix of its corresponding training sample one after another.

$$\Gamma = [\Gamma_1 \Gamma_2 \dots \Gamma_M] \quad (1)$$

And its size is (Q x M) where M is the number of training samples.

Now column wise mean of this training matrix is found out.

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (2)$$

Its size is Q x 1.

Now difference of each approximation matrix is calculated from just found mean image vector.

$$\Phi = \Gamma - \Psi \quad (3)$$

Each of mean subtracted approximation matrix are then appended column wise to form difference matrix.

$$A = [\Phi_1 \Phi_2 \Phi_3 \Phi_4 \dots \Phi_M] \quad (4)$$

Its size will be Q x M. Here also, M is the number of training samples.

Covariance of difference matrix is then found, which its multiplication with its own transpose is

$$X = AA' \text{ and its size is } Q \times Q \quad (5)$$

The size of this covariance matrix is very big and will take lots of time so cost of recognition will get increased tremendously. Hence we apply PCA (Principle Component Analysis) here to calculate the Eigen vectors of size M x M.

According to PCA,  $A'$  and  $A$  are multiplied to get a square matrix ( $Y = A'A$ ) from where the Eigen vectors ( $V_i$ ) are calculated. Where  $i = 1, 2, 3, \dots, M$

Since most of the generalization power is contained in the first few eigenvectors. Around 40% of the total eigenvectors have 85 – 90 % of the total generalization power [21]. So instead of using all the M numbers of the Eigen vectors,  $M' \leq M$  numbers of the Eigen vectors are sufficient to show its effect.

The Eigen vector of X ( $U_j$ ) is derived from  $V_j$  for weight calculation.

$$U_j = AV_j \text{ where } j = 1, 2, 3, \dots, M' \quad (6)$$

Each of mean subtracted training approximation matrix are mapped to Eigen space to calculate the feature vector of each training sample using the formula

$$\Omega_i = \Phi_i U_j \text{ where } j = 1, 2, 3, \dots, M' \text{ and } i = 1, 2, 3, \dots, M \quad (7)$$

#### 3.4 Mapping to Fisher Space

The aim of this phase is to maximize the scatter between different classes and minimize the scatter between the input data in the same class.

Suppose in the training database, there are  $C$  individual person classes having  $q_s$  number of training images in each of the class  $s$ , such that total number of facial images  $M = C \times q_s$

For all training samples of all classes there are two measures (1) within-class scatter matrix ( $S_w$ ) and (2) between-class scatter matrix ( $S_b$ ). These two matrices will be calculated based on Eigen feature.

The Eigen space class mean is

$$\Theta_s = 1/q_s \sum_{k=1}^{q_s} \Omega_k \quad \text{where } s=1, 2, \dots, C \quad (8)$$

The arithmetic average of all the Eigen space mapped training feature vectors

$$\Theta = 1/M \{\Omega_1 + \Omega_2 + \Omega_3 + \dots, \Omega_M\} \quad (9)$$

Within class scatter matrix

$$S_w = \sum_{s=1}^C \sum_{k=1}^{q_s} (\Omega_k - \Theta_s)(\Omega_k - \Theta_s)' \quad (10)$$

Between class scatter matrix

$$S_b = \sum_{s=1}^C (\Theta_s - \Theta)(\Theta_s - \Theta)' \quad (11)$$

For good fisher feature extraction we have to maximize the ratio of (determinant of ( $S_b$ )): determinant of ( $S_w$ ))

So we take 40% of most significant Eigen vectors ( $L$ ) of  $S_w^{-1}S_b$  [21] to calculate weight vector where  $S_w^{-1}$  is inverse of  $S_w$ .

Now the Eigen space class mean will map to fisher space using the formula

$$\Pi_s = \Theta_s L \quad \text{where } s=1, 2, \dots, C \quad (12)$$

### 3.5 Testing

The second level approximation matrix of test image is of size ( $R \times C$ ). It is also converted into a single column vector,  $\Gamma_T$ , to represent in image space. Its size will be ( $Q \times 1$ ).

The difference of test approximation matrix from the mean is

$$\Phi_T = \Gamma_T - \Psi \quad (13)$$

Each of mean subtracted test approximation matrix is now projected to Eigen space using the formula

$$\Omega_T = \Phi_T U_j \quad \text{where } j = 1, 2, 3, \dots, M' \quad (14)$$

Now each Eigen features of test image are projected to fisher space using the formula

$$\Pi_T = \Omega_T L \quad (15)$$

### 3.6 Classification

The distance between test feature  $\Pi_T$  and average training class feature  $\Pi_s$  (where  $s$  represents the number of classes) is measured using Manhattan distance. The test sample  $t$  of feature value  $\Pi_T$  is classified to class  $k$  if the absolute difference  $D_s, t$

between test sample  $t$  and  $s$ -th training class is minimum for  $s=k$  which is

$$t \rightarrow k, \quad \text{if } D_s, t = |\Pi_T - \Pi_s| \text{ is minimum for } s=k \quad (16)$$

Further we have fixed a distance threshold to handle unknown faces by calculating the maximum distance between two training class using the formula.

$Th = \max(|\Pi_a - \Pi_b|)/NC$  where  $a, b=1, 2, 3, \dots, C$  and  $NC$  is number of classes. If  $(D_s, t)/NC \geq Th$  for all values of  $s$  then the face is an unknown face else known face and the class will be determined using the formula (16).

## 4. EXPERIMENTAL RESULTS

All simulation based experiment reported here are performed using the YALE face database [23], MUCT face database [24] and Japanese female facial expression database (JAFFE) [25].

For YALE database there are 165 images of 15 subjects each having 11 images. Among the 11 images of each class 6 images are used for the feature extraction and 5 for testing. Each of the images is in BMP format.

For MUCT face database we take 750 images of 50 subjects each having 15 images. Among the 15 images of each class 10 images are used for the feature extraction and 5 for testing. Each of the images is in JPEG format.

For JAFFE face database we take 200 images of 10 subjects each having 20 images. Among the 20 images of each class 10 images are used for the feature extraction and 10 for testing. Each of the images is in TIFF format.

### A. Training Phase



Figure 2. Sample training images for YALE dataset



Figure 3. Sample training images for MUCT dataset



Figure 4. Sample training images for JAFFE dataset

The following figure shows the feature values of YALE database. Up to 14 most significant discriminant features are plotted below by taking their average.

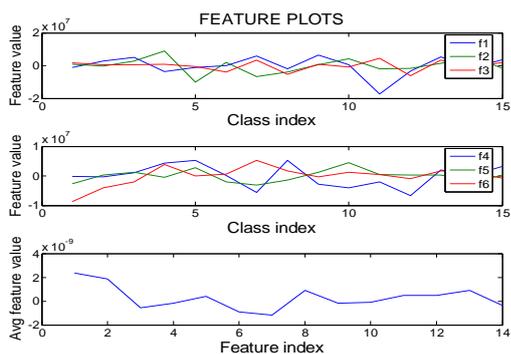


Figure 5. Feature Values of Training phase of YALE

The following figure shows the feature values of MUCT database. Up to 20(out of 80) most significant discriminant features are plotted below by taking their average.

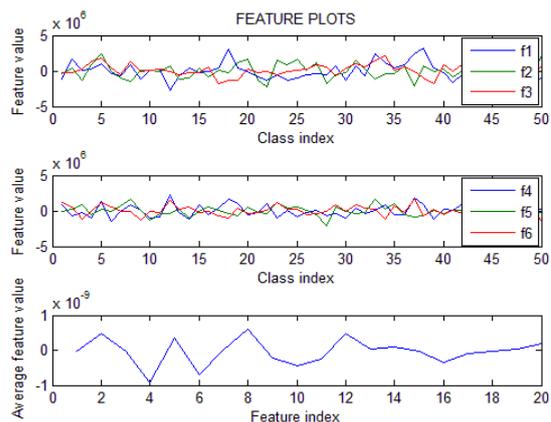


Figure 6. Feature Values of Training phase of MUCT

The following figure shows the feature values of JAFFE database. Up to 16 most significant discriminant features are plotted below by taking their average.

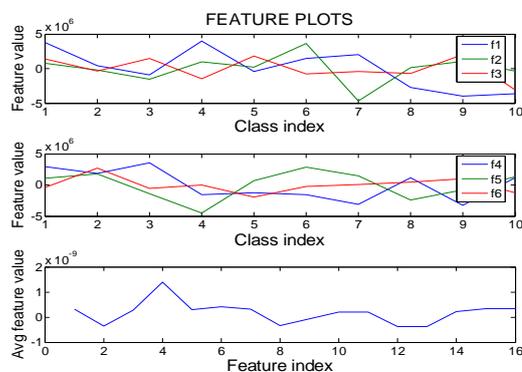


Figure 7. Feature Values of Training phase of JAFFE

## B. Testing Phase



Figure 8. Sample testing images for YALE dataset

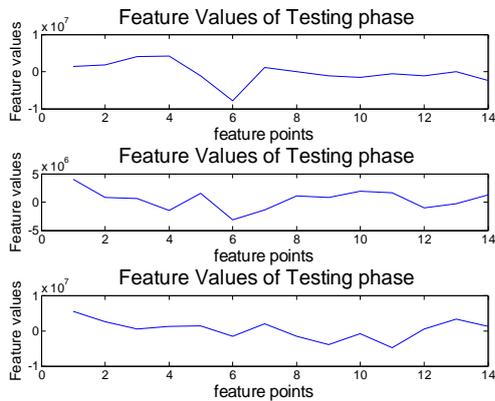


**Figure 9. Sample testing images for MUCT dataset**



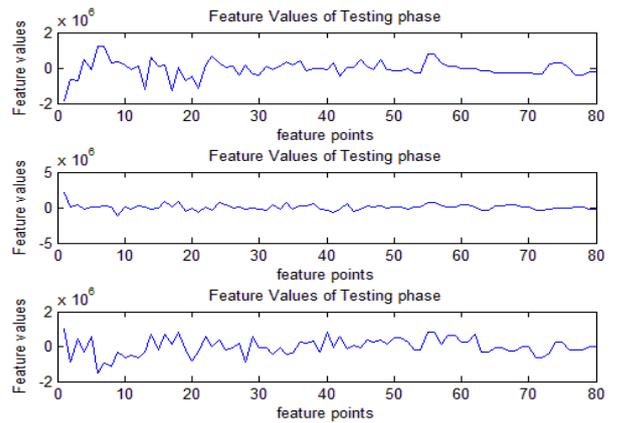
**Figure 10. Sample testing images for JAFFE dataset**

The following figure shows the feature values of 3 test samples which belong to class 1, 2 and 3 respectively for YALE dataset.



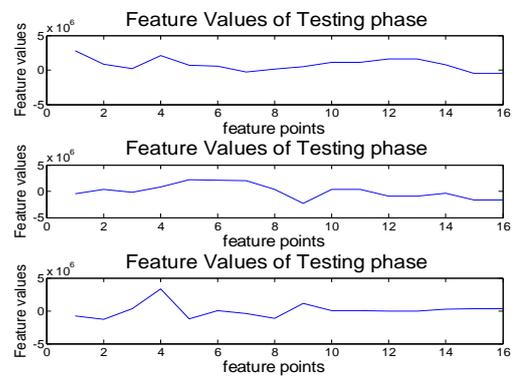
**Figure 11. Feature values of testing phase for YALE**

The following figure shows the feature values of 3 test samples which belong to class 1, 2 and 3 respectively for MUCT dataset.



**Figure 12. Feature values of testing phase for MUCT**

The following figure shows the feature values of 3 test samples which belong to class 1, 2 and 3 respectively for JAFFE dataset.

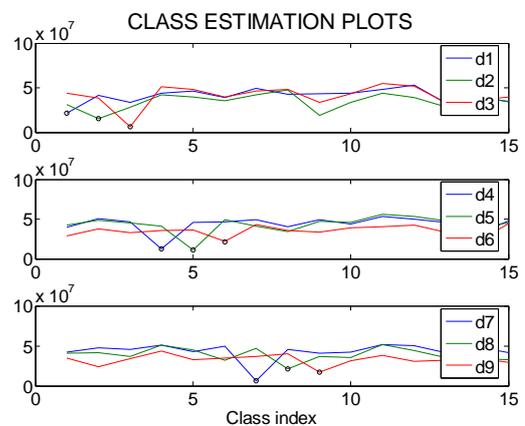


**Figure 13. Feature values of testing phase of JAFFE**

### C. Class Estimation

In this phase the class estimation plot is explained. The difference plot of the test sample with all the training classes. The minimum difference corresponds to the estimated class of the test sample.

We take test samples from first 9 classes for YALE, JAFFE and MUCT dataset to plot the class estimation.



**Figure 14. Class estimation plot For YALE dataset**

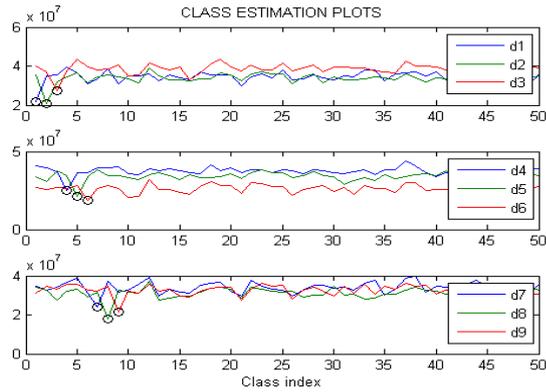


Figure 15. Class estimation plot For MUCT dataset

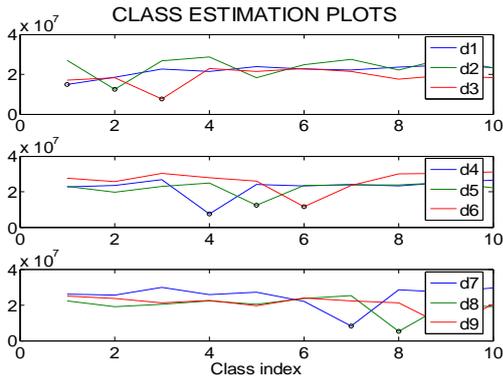


Figure 16. Class estimation plot For JAFFE dataset

In the following figure we have plotted recognition rate with respect to the number of feature values for JAFFE and YALE dataset. It shows the more number of features, the higher recognition rate.

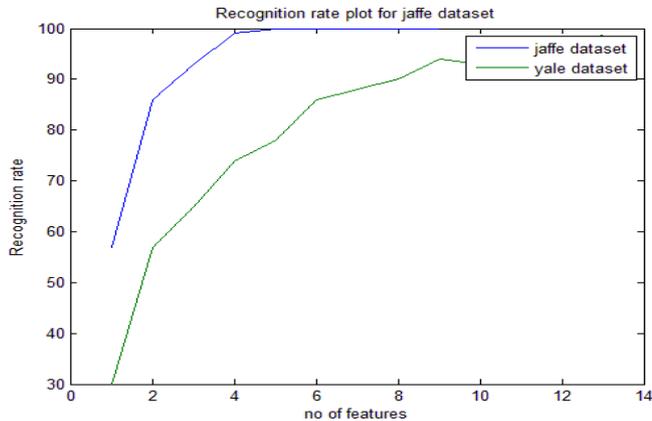


Figure 17. Recognition rate plot for JAFFE and YALE dataset

Below the Time Complexity is calculated for all three datasets.

Dataset	Time Complexity(in second)
JAFFE	5.45
YALE	4.85
MUCT	164.21

Figure 18. Time complexity for all three datasets

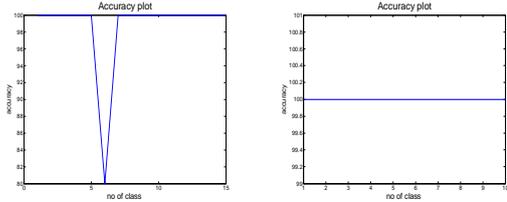


Figure 19. Accuracy plot For YALE and JAFFE dataset

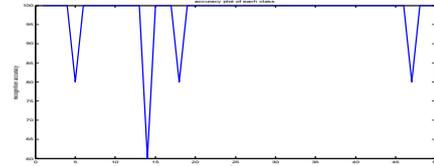


Figure 20. Accuracy plot For MUCT dataset

#### D. Unknown Face Handling

From the experiment we have determined that the threshold will be more accurate to eliminate unknown faces if the number of classes is high. For this reason we have used MUCT dataset in Training phase. In testing we collect 50 different person's face from MUCT dataset and 25 different faces from YALE and JAFFE dataset. The following figure illustrates the result. In x-axis test samples and in y-axis Minimum class difference.

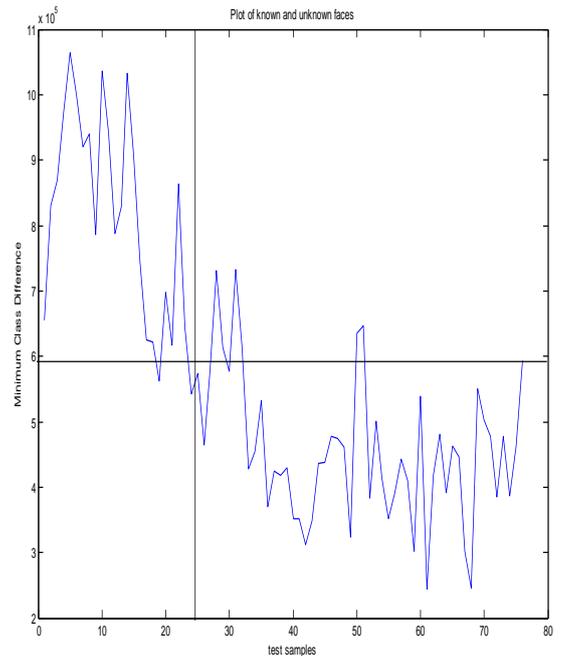


Figure 21. Plot to recognize known and unknown faces. First 25 faces are unknown and next 50 faces are known

In our experiment out of 50 known faces 6 faces are miss classified as unknown and out of 25 unknown faces 3 faces are miss classified as known. False Rejection Rate =12% , False Acceptance Rate=12%.

## 5. ANALYSIS

The recognition accuracy for discrimination of the face image Of each class is done successfully and with satisfactory accuracy.

In current work we have incorporated Histogram Equalization to handle various illumination conditions on human face and tested on YALE, MUCT and JAFFE dataset which contains apparent changes of illumination.

The main purpose of PCA is to reduce the dimensionality of the raw data while retaining as much as possible the variation present in the dataset [22]. But the drawback is, it maximizes within class scatter also, so Fisher analysis reduces the scattering of projected data since it is a class specific method and masks portions of faces which are highly variable[6]. So we have incorporated both PCA and Fisher analysis in current work to handle various expressions on human face and tested on YALE, MUCT and JAFFE dataset which contains apparent changes of facial expressions.

The following figure shows that the feature value is not distracted for various facial expressions of a person but two different persons are discriminated successfully.

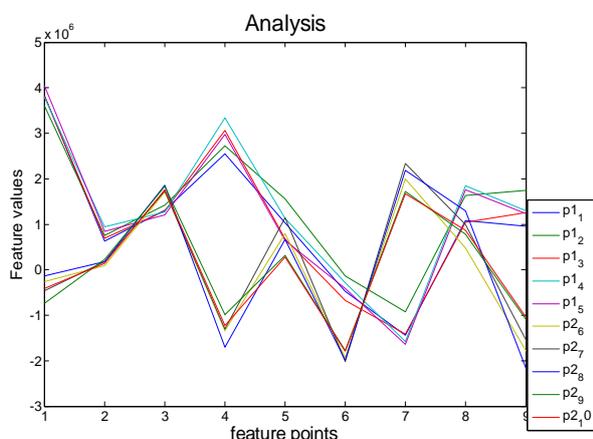


Figure 22. Features for 10 samples of 2 persons with various Facial Expressions of JAFFE dataset

We analyzed the most similar work we could find in [4] and [13] and tested with the same above and can see that the most compact feature for various facial expressions which we are getting from our current work.

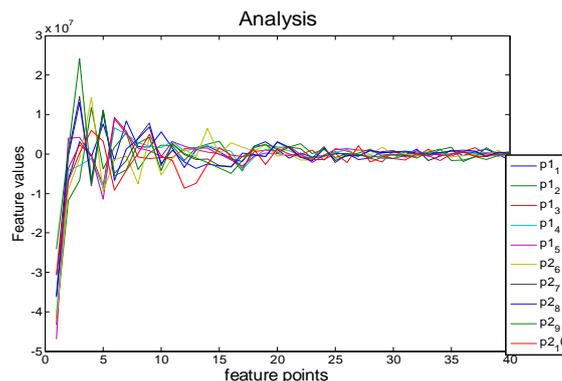


Figure 23. For [4]

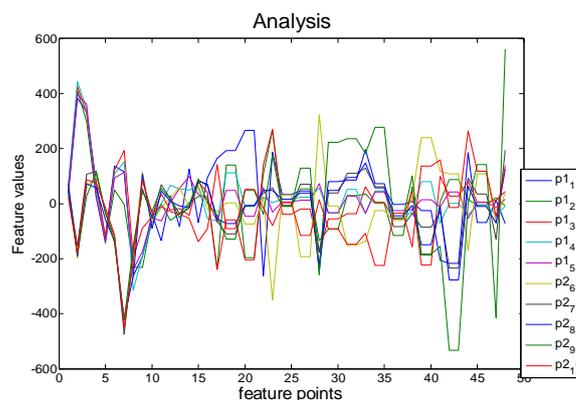


Figure 24. For [13]

[4] and [13] are not suitable where various illumination conditions are present. So in our current work HE is performed and recognition accuracy is improved. Also we have proposed a mechanism to deal with unknown faces in our work. Moreover a good response time enables recognition in real time applications. The following table compares the recognition accuracy.

DATASET	JAFFE	YALE	MUCT
Previous work in [4]	86%	80%	10.80%
Previous work in [13]	99%	81.33%	60%
Current work	100%	98.66%	98%

Figure 25. Accuracy table

## 6. CONCLUSIONS AND FUTURE SCOPES

In this paper we have proposed an algorithm for human face recognition which can handle various facial expressions and illumination condition. We have used wavelet transform, Eigen and Fisher analysis. Initially Histogram Equalization is performed to handle illumination condition and then wavelet transform is performed to reduce the time complexity. The transformed approximation matrix is then mapped to Eigen space and further mapped to fisher space for feature extraction.

Classification is done using Manhattan distance and a distance threshold is calculated to distinguish between known and unknown faces. A critical issue using LDA is the Very Small Sample Size (VSSS) problem. This problem arises when there is a very small number of training samples but the dimension of the feature space is large. This means that the within class scatter matrix would tend to be a singular matrix and the execution of LDA may encounter computational difficulty. In our experiment when we are taking only a single sample to train the system then it encounters a problem and can't be able to distinguish persons. This work can be extended further by incorporating methods like Adaptive Discriminant Learning (ADL), Singular Value Decomposition (SVD) and Two Dimensional LDA (2DLDA). In ADL a generic discriminant model is adapted for identification. SVD and 2DLDA can also be applied to deal with this problem.

## 7. REFERENCES

- [1] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, "Face recognition: literature survey," *ACM Computing Surveys*, pp. 399–458, 2003.
- [2] L. Wiskott, J.M.Fellous, N.Kunger, and C. Vonder Malsburg, "Face Recognition by elastic bunch graph matching," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779, July 1997.
- [3] W. W Bledsoe, "The model method in facial recognition," *Panoramic Research Inc., Palo Alto, CA, Rep. PRI. 15*, August 1966.
- [4] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuro-science*, vol. 3, no. 1, pp. 72-86, 1991.
- [5] Zhao W., Chellappa R., Krishnaswamy A., "Discriminant Analysis Of Principal Component For Face Recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol 8, pp. 336-341, April 1998.
- [6] Peter N. Belhumeur, Jo~ao P. Hespanha, David J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997.
- [7] J. Wang and H. Yang, "Face Detection Based on Template Matching and 2DPCA Algorithm," *IEEE Transactions on Congress on Image and Signal Processing*, pp. 575-579, May 2008.
- [8] R. Gottumukkal and V.K. Ansri, "An improved face recognition technique based on Modular PCA approach," *Pattern Recognition letters*, vol 25, pp. 429-436, 2004.
- [9] T. D. Diwan, S. Mishra and R. Venma, "ICA Based Face Authentication System," *International Journal of Computer Science and Telecommunications*, vol. 3, no. 4, pp. 103-107, April 2012.
- [10] H. Intiaz and S. A. Fattah, "A Wavelet-Domain Local Feature Selection Scheme for Face Recognition," *IEEE International Conference on Communications and Signal Processing*, pp. 448-451, September 2011.
- [11] S. K. Jawad, "Design a Facial Recognition System Using Multilayer Perception and Probabilistic Neural Networks Based Geometrics 3D Facial," *European Journal of Scientific Research*, vol. 60, no. 1, pp. 95- 104, 2011.
- [12] X. Y. Jing and D. Zhang, "A face and palm print recognition approach based on discriminant DCT feature extraction," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 34, pp. 2405–2415, 2004.
- [13] P. Marasamy and S.Sumathi, "Automatic Recognition and Analysis of Human Faces and Facial Expression by LDA Using Wavelet Transform," *2012 International Conference on Computer Communication and Informatics*, pp. 1-4, January 2012.
- [14] Yung-Mao Lu, Bin-Yih Liao, Jeng-Shyang Pan, "A Face Recognition Algorithm Decreasing the Effect of Illumination," *2008. IJHMSP '08 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 378-381, August 2008.
- [15] C.MageshKumar, R.Thiyagarajan, S.P.Natarajan, G.Sainarayanan, "Gabor features and LDA based face recognition with ANN classifier," *2011 International Conference on Emerging Trends in Electrical and Computer Technology*, pp. 831-836, March 2011.
- [16] Jyothi S. Nayak, Indiramma M, Nagarathna N, "Modeling self-Principal Component Analysis for age invariant face recognition," *2012 IEEE International Conference on Computational Intelligence and Computing Research*, pp. 1-5, December 2012.
- [17] Weihong Deng, Jiani Hu, "Transform-Invariant PCA: A Unified Approach to Fully Automatic Face Alignment, Representation, and Recognition," *IEEE Transactions on Pattern analysis and Machine Intelligence*, Vol. 36, No. 6, pp. 1275-1284, June 2014.
- [18] S.S.Meher, Pallavi maben, "Face recognition and facial expression identification using PCA," *2014 IEEE International Advance Computing Conference*, pp. 1093-1098, February 2014.
- [19] M. Abdulrahman, Tajuddeen R. Gwadabe, A. Eleyan, F.J. Abdu, "Gabor wavelet transform based facial expression recognition using PCA and LBP," *2014 22nd Signal Processing and Communications Applications Conference*, pp. 2265-2268, April 2014.
- [20] Abhijit Das, "Face recognition in reduced Eigen-plane," *2012 International Conference on Communications, Devices and Intelligent Systems*, pp. 620-623, December 2012.
- [21] Ahmet Bahtiyar Gul, "Holistic Face Recognition by Dimension Reduction", Ph.D. Thesis, M.Sc., EEE, Middle East Technical University, 2003.
- [22] C. Liu, H. Wechsler, "Independent Component Analysis of Gabor Features for Face Recognition", *IEEE Transactions on Neural Networks*, Vol. 14, pp. 919-928, 2003.
- [23] Yale Face Dataset, Available ([www.cvc.yale.edu](http://www.cvc.yale.edu)).
- [24] MUCT Face Dataset, Available (<https://code.google.com/p/muct/downloads/list>).
- [25] JAFFE Face Dataset, Available (<http://www.kasrl.org/jaffe.html>)