FACE RECOGNITION SYSTEM USING PCA, LDA, KERNEL PCA AND KERNEL LDA

Article · January 2019

Esra Jasem
University of Basrah

International Journal of Computer Science Engineering
and Information Technology Research (IJCSEITR)
ISSN(P): 2249-6831; ISSN(E): 2249-7943
Vol. 6, Issue 5, Oct 2016, 9-20

FACE RECOGNITION SYSTEM USING PCA, LDA, KERNEL PCA AND KERNEL LDA

ESRA JASEM HARFASH

Department of Computer Science, Collage of Science, Basrah University, Basrah, Iraq

ABSTRACT

Automatic Face recognition (AFR) is an important research topic in pattern recognition field where there are many subjects in this area. In this research, we focused on the problem of features extraction by using: PCA and LDA and kernel-based methods are kernel PCA (KPCA) and kernel LDA(KDA). We are trying here to get to that these methods have the effect of significant impact in the extracted important and discriminative features and therefore its impact on the results of face recognition. We were applied these methods to the image directly, and also applied it to the DCT and FFT coefficients .These features extraction methods have shown different performance effects when applied to different information that extracted from the image. In general the experimental results show that PCA/LDA/KPCA and KDA are able to extract significant features and greatly reduce the dimension of images, and then yield good accuracies in face recognition.

KEYWORDS: Automatic Face Recognition, Principal Component Analysis PCA, Linear Discriminant Analysis LDA, FISHER Discriminant, Features Selection

Original Article

Received: Aug 19, 2016; **Accepted:** Sep 06, 2016; **Published:** Sep 12, 2016; **Paper Id.:** IJCSEITROCT20162

INTRODUCTION

In recent years, Automatic face recognition is one of the important technology that is being used for the last years, and has attracted much attention and its research has rapidly expanded by not only engineers but also neuroscientists, since it has many potential applications in computer vision communication and automatic access control system[1].The face recognition system is a computer application capable of identifying or verifying a person from a digital image or a video frame from a video source. One of the ways to do this is by comparing selected facial features from the image and a facial database. It is typically used in security systems and can be compared to other biometrics such as fingerprint or eye iris recognition systems. Recently, it has also become popular as a commercial identification and marketing tool [2] [3].

Over the last few decades many techniques have been proposed for face recognition. All of these a techniques to human face recognition can be divided into two strategies: geometrical features (This technique involves computation of a set of geometrical features such as nose width and length, mouth position and chin shape. This set of features is then matched with the features of known individuals by calculating the Euclidean distance [4], and template matching(here we are not trying to classify an image as a 'face' or 'non-face' but are trying to recognize a face by extract the whole facial regions :matrix of pixels, and compare these with the stored images of known individuals. Where the feature based strategy may offer higher recognition speed and smaller memory requirements, template based techniques offer superior recognition accuracy)[5].

However, for many patterns sets with a large number of features and a limited number of observations, such as bioinformatics data, usually many features are not useful for producing a desired learning result and the limited observations may lead the learning algorithm to over fit to the noise. Then the Solution to a number of problems in Pattern Recognition can be achieved by choosing a better feature space. The appropriate selected features should be robust against geometry distortion caused by viewpoint change and informative so that they capture the uniqueness of an object. The benefit of using such features include shorter inverted list thus memory efficient, higher relevance for the result candidate set and faster recognition by accelerated ranking [6].

This paper shows some popular feature selection methods are PCA/LDA/KPCA and KDA and its effective in Automatic Face recognition depending on face94 and face95 database. To find out the efficiency of each method were applied as following:1) directly on images, 2) on DCT images features and 3) on FFT images features.

Face Recognition System

The Face Recognition system consist of important sequence of tasks as in figure 1 .In our work ,we are followed these processes steps to achieve the face recognition problem according to face94 and face95 database.

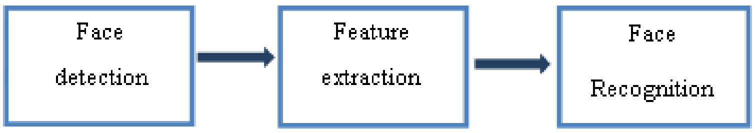

Figure 1: The General Face Recognition System

Face Detection

This process is considered one of the key process, which can be defined as the process of determine and extract the faces from the input images or video image. The methods that is used to implement this task may be involves segmentation, extraction, and verification of faces and possibly facial features from an uncontrolled background [7].

We are depended on Viola-Jones algorithm to detect the face part. This algorithm should be capable of functioning in an unconstrained environment, this meaning that it should detect all visible faces in any conceivable image [8] .Figure 2 two images before and after the face detection process.

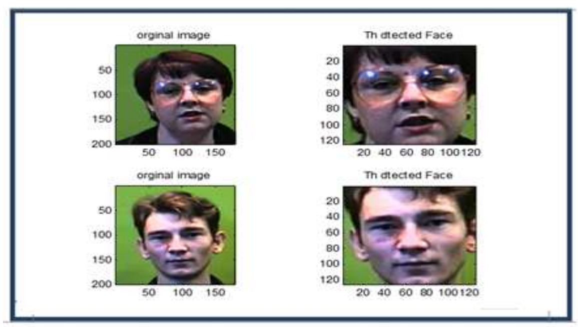

Figure 2: A Successful Face Detection

Features Extraction

This task is defined as the process of extraction the facial features from the input images. The features could be face regions, variations, angles or measures, which can be human relevant (e.g. eyes spacing) or not. This phase has other applications like facial feature tracking or emotion recognition. [10].

We are focused in this process on PCA, LDA, KPCA and KDA of extracted important information from the face, as we tested these methods, as shown in Figure 3, and as follows:

- Apply directly on pixel image intensities
- Apply on DCT Coefficients of image ,
- And apply on FFT Coefficients of image

The above tests were applied to the training data after taking the overall averages, also been applied in the case of taking the average of each class data.

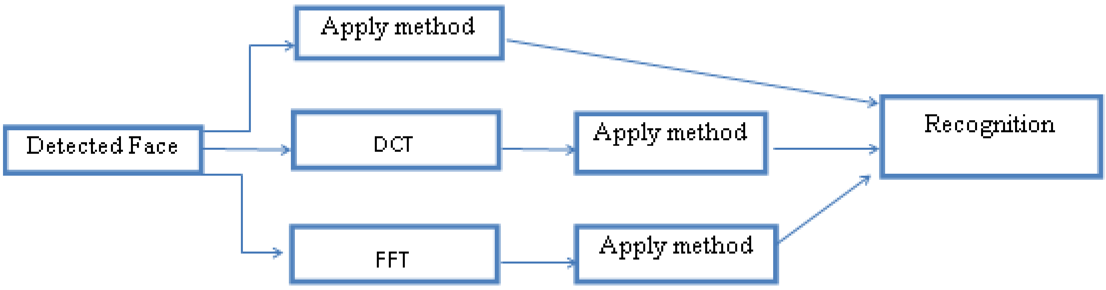

Figure 3: The General Outline of the Feature Extraction.
Method is one of (PCA, LDA, KPCA and KDA)

The following the methods that was depended in this research:

- **DCT (Discrete Cosine Transform):**

In case of images, this transformation is used to map the spatial variation into uncorrelated frequency variation [10]. We are used the two-dimensional DCT to extract the important coefficients in faces. The two-dimensional DCT is defined as:

$$F(x,y) = W(x)W(y)\sum_{i=0}^{M-1}\sum_{j=0}^{M-1} Y(i,j)cos\left[\frac{\pi(2i+1)x}{2M}\right] cos\left[\frac{\pi(2j+1)j}{2M}\right] \quad (1)$$

For x,y= 0,1,2,…,M−1 and $W(x)$ and $\alpha(y)$ are defined :

$$W(k) = \begin{cases} \sqrt{\frac{1}{M}} for\ k = 0 \\ \sqrt{\frac{2}{M}} for\ k \neq 0 \end{cases} \quad (2)$$

We are followed the following procedure to calculate DCT coefficients of images:

- Convert the images MxN to a gray images MxN, and normalize the resulted gray image .
- Apply DCT as in equation 1 on the resulted images from step1, the result is MxN DCT Coefficients as show in figure 4
- We are selected the first 10x10 upper left corner of DCT matrix
- Scan the resulted matrix in step 3 in Zigzag scan.
- After the scan ,select 100 features to represent the each image

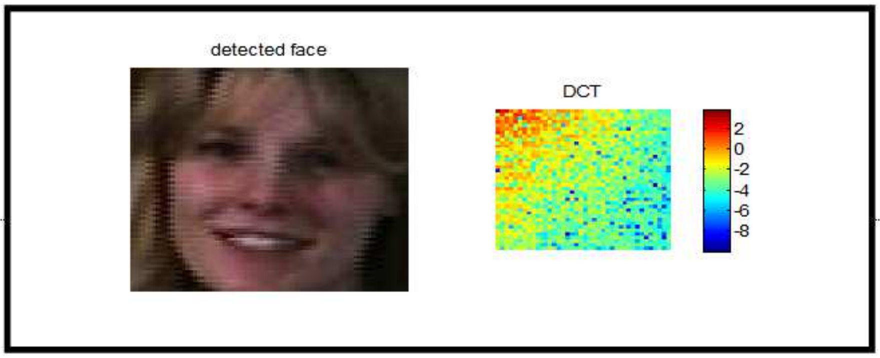

Figure 4: The DCT Coefficients

- **Fast Fourier Transform**

The Fourier Transform is an important image processing tool. This Transformation is used if we want to access the geometric characteristics of a spatial domain image. The Fast Fourier Transform (FFT) is an efficient and fast algorithm to compute the discrete Fourier transform (DFT) and its inverse[11]. The DFT is given in the following equation**:**

$$F(x,y) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} Y(i,j) e^{-i2\pi\left(\frac{xi}{M}+\frac{yj}{N}\right)} \tag{3}$$

Where Y(i,j) is the image in the spatial domain and the each point F(x,y) in the Fourier space. The following the steps are followed to calculate FFT of images:

- Convert the images to gray images and then normalize it.
- Apply FFT according to the equation 3 .In most implementations the Fourier image is shifted that the value (i.e. the image mean) F(0,0) is displayed in the center of the image.
- Use the abs and then log functions :abs (log(FFT) to compute the magnitude of the combined components. See figure 5.
- We know that the second half of FFT carry no useful and duplicated information, so we can half the data to treat.

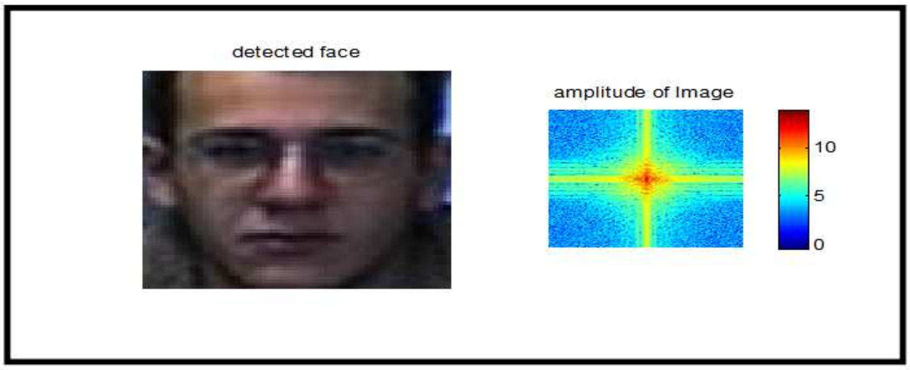

Figure 5: The Resulted FFT Coefficients

- **The PCA Technique**

Principal component analysis PCA is standard technique used in statistical pattern recognition and signal processing for data reduction and Feature extraction [12].It is the way to find patterns in the data, and expressing the data as a way to find similarities and differences between them. The analyzing patterns in a high-dimension data may be difficult, then the PCA represent a powerful method for dealing with such data. The other benefit of PCA is not loss the important information through identifying the patterns by reducing the number of dimensions.

The PCA face recognition finds eigenvectors, these eigenvectors are called eigenfaces that represent the global feature of the training images.[13]. For Given training set Y1, Y2,.. ., Y_N:

- Compute the mean vector

a. $Me = \frac{1}{N}\sum_{i=1}^{N} Y_i$ (4)

- Compute the *covariance matrix*

b. $C = \frac{1}{N}\sum_{i=1}^{N}(Y_i - Me)(Y_i - Me)^t$ (5)

- Compute the eigenvalue/eigenvector pairs (λi, ui) of C, $1 \leq i \leq N$, where $\lambda 1 \geq \lambda 2 \geq \cdot\cdot\cdot \geq \lambda_N$.
- Compute the first k principal components $z_j^{(i)} = Yj^t ui$, for each observation Y_j, $1 \leq j \leq n$, along the direction ui, i= 1, 2, · · ·, k. where $z_j^{(i)}$ is the eigenface.

In the Figure 6 show the first 8eigenface that corresponding to the largest eigenvalue.

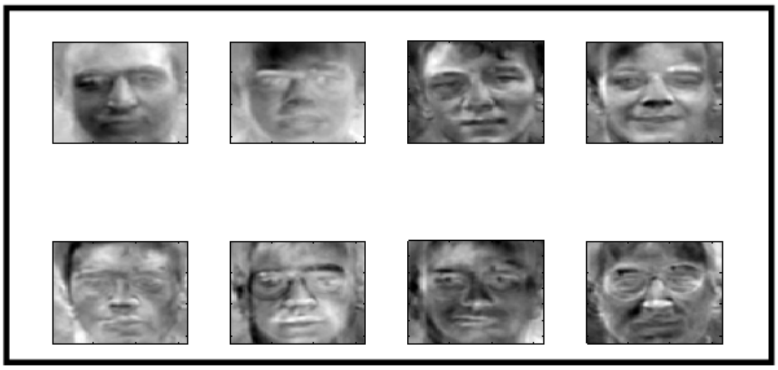

Figure 6: Eigen Faces from one of the 8 Training Sets with largest Eigenvalues

- **The KPCA technique**

The kernel principal component analysis is the nonlinear version of PCA that is constructed by using a kernel function is. By using a nonlinear mapping, the data set can be mapped into a higher dimensional feature space *F*. For certain feature spaces *F* there is a function for computing scalar products in feature spaces .Given a set of centered data $Y_k \in R^N$ [14][15]:

$\sum_{k=1}^{M} \Phi(Y_k) = 0$ (6)

PCA diagonalized the covariance matrix in *F:*

$C = \frac{1}{M}\sum_{i=1}^{M} \Phi(Y_i)\, \Phi(Y_i)^T$ (7)

To find the eigenvalues $\lambda \geq 0$ and eigenvectors satisfying $\lambda \mathbf{v} = C\mathbf{v}$ All solutions v with $\lambda \neq 0$ must lie in the span of Φ (Y1), Φ (Y2),…, Φ (Y*M*). Then there exist coefficients α *i* (*i*= 1, 2, …, *M*) such that :

$$v = \sum_{i=1}^{M} \alpha_i \, \Phi(Y_i) \tag{8}$$

By defining the kernel function as

$$k(Y_i, Yj) = \Phi(Y_i).\, \Phi(Y_j) \tag{9}$$

and $K_{ij}=(\Phi(Y_i).\, \Phi(Y_j))$ i=1….M and j=1….M Then :

$$M\, \lambda K\alpha = K^2\, \alpha \tag{10}$$

where α is the column vector with entries α 1,…, α_M. The eigenvalue problem is solved$M\lambda\boldsymbol{\alpha} = K\boldsymbol{\alpha}$.

For the principal component extraction, the projections onto the eigenvectors v^kin *F* are needed.

$$(v^k.\, \Phi(Y)) = \sum_{i=1}^{M} \alpha_i^k \, (\Phi(Y_i).\, \Phi(Y)) \tag{11}$$

Where, Y as a test point and Φ(Y) is its image in *F*. With this method (as in PCA) we are:

- Find the average of training data
- Compute the eigenvector, and eigenvalue, see figure 7
- Then projections test images onto the resulted eigenvectors

Figure 7: First 8 Eigenvectors with Highest Eigenvalues (KPCA)

- **The LDA Technique:**

Linear discriminant analysis (LDA) is used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier or, more commonly, for dimensionality reduction before later classification. LDA explicitly attempts to model the difference between the classes of data. Suppose there are *N* classes and A V*i* be the mean vector of class *i*, *i*=1, 2,.. ,*N*, and of the *Mi* is the number of samples in class *i*, *i*=1, 2,.. ,*N*, then the total number of samples is[16][17]:

$$M = \sum_{i=0}^{N} M_i \tag{12}$$

And Within-class scatter matrix:

$$S_w = \sum_{i=1}^{N}\sum_{j=1}^{M_i}(y_j - \mu_i)(y_{j-} - \mu_i)^T \quad (13)$$

Between-class scatter matrix:

$$S_b = \sum_{i=1}^{N}(\mu_i - \mu)(\mu_{i-} - \mu)^T \quad (14)$$

$$\mu = 1/N \sum_{i=1}^{N} \mu_i \quad (15)$$

The linear transformation is given by a matrix *U* whose columns are the eigenvectors of $S^{-1}_w S_b$ (called *Fisherfaces*).

$$\begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ b_K \end{bmatrix} = \begin{bmatrix} u_1^T \\ u_2^T \\ . \\ . \\ u_K^T \end{bmatrix}(x - \mu) = U^T(x - \mu) \quad (16)$$

The eigenvectors are solutions of the *generalized eigenvector problem*: $S_B u_k = \lambda_k S_w u_k$ .There are at most *N*-1 non-zero generalized eigenvectors (i.e., *K*<*N*).We are Followed the following steps to compute the LDA eigenvector:

- Compute the average of training data
- Computer the highest LDA eigenvector, see figure 8
- Project Image Y_i to a linear subspace $\boldsymbol{E_i} = \boldsymbol{V_t}\, \boldsymbol{Y_i}$, where $\boldsymbol{V}$ is called projection matrix(*v1*, *v2* ,…, *vN-1* be the corresponding eigenvectors)
- Classify by nearest neighbor

Figure 8: First 8 Eigenvectors with Highest Associated Eigenvalues (LDA)

- **The Kernel LDA Technique**

Kernel Methods are a new class of pattern analysis algorithms which can operate on very general types of data and can detect very general types of relations. We Can solve any problem in non-linear version of the Fisher's LDA, and to do so, we would define between-class and within-class covariance matrices in kernel space F ,we need to find eigenvalues λ and eigenvectorW^{Φ} for c class and M number of samples in each class, the new feature space[18][19]:

$$J(w) = \frac{w^T s_B^{\Phi} w}{w^T s_W^{\Phi} w} \quad (17)$$

Where

$$S_w = \sum_{i=1}^{c} \sum_{j=1}^{M_i} (\Phi(Y_j^i) - \mu_i^\Phi)\left(\Phi(Y_j^i) - \mu_i^\Phi\right)^T \quad (18)$$

Between-class scatter matrix:

$$S_B^\Phi = \sum_{i=1}^{c} M\left(\mu_i^\Phi - \mu^\Phi\right)\left(\mu_i^\Phi - \mu^\Phi\right)^T \quad (19)$$

$$\mu_i = \frac{1}{M} \sum_{j=1}^{M} y_j^i \quad (20)$$

wwill have an expansion of the form:

$$w^\Phi = \sum_{i=1}^{c} \sum_{j=1}^{M} \alpha_j^i \, \Phi(Y_j^i) \quad (21)$$

$$w_{OPT}^\Phi = argmax_{w^\Phi} \frac{|(w^\Phi)^T S_B^\Phi w^\Phi|}{|(w^\Phi)^T S_W^\Phi w^\Phi|} = [w_1^\Phi \ldots\ldots . w_c^\Phi] \quad (22)$$

We can project $\Phi(Y)$ to alower dimensional space spanned by the eigenvectors W^Φ.The following is the algorithm that followed to calculate KDA and classification:

- Compute the average of training data
- Computer the highest KDA eigenvector, see figure 9.
- Project test Image Y_i to a linear subspace to corresponding eigenvectors
- Classify by nearest neighbor

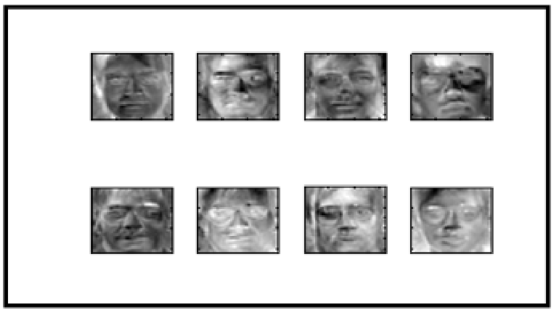

Figure 9: First 8 Eigenvectors with Highest Associated Eigenvalues (KDA)

Face Recognition Techniques

In our research we have adopted Euclidian Distance and KNN methods at the recognition phase, the following explanation of these two methods**:**

- **Euclidian Distance Method:**

In classification problems the distance measures is to determine the similarity or dissimilarity between any pair of objects. It is useful to denote the distance between two instances x_i and x_j as: d (x_i,x_j). By using the Euclidean distance metric, the distance between each data object calculate using the equation below[20]:

$$Dist = \sqrt{\sum_{k=1}^{N} (X_{ik} - X_{jk})^2} \quad (23)$$

The minimum distance mean the more similarity

- **K Nearest Neighbour Method**

The way in which the algorithm decides which of the points from the training set are similar enough to be considered when choosing the class to predict for a new observation is to pick the k closest data points to the new observation, and to take the most common class among these. The algorithm can be summarized as [21][22]:

- A positive integer k is specified, along with a new sample
- We select the k entries in our database which are closest to the new sample
- We find the most common classification of these entries
- This is the classification we give to the new sample

EXPERIMENTAL RESULTS

The performance of the proposed method has been evaluated on the face94 and face95 database. Many experiments have been carried out on these features selection methods to test its performance. In the test phase, the faces in testing set images are presented to the features selection methods to effect the recognition. In the following tables shows the recognition rate and complexity time of train data for PCA/LDA/KPCA /KDA .

In tables 1 shows the results of the recognition accuracy in the case of apply selection features methods of directly on the images training data, as in the this tables show the results with the averages of all face images, either note where results were taken Average per class.

Table 1: Accuracies of Direct Apply

Direct	Average of All Training set		Average of each Classes	
	DIST	KNN	DIST	KNN
PCA	95%	95%	89%	88.5%
LDA	93.3%	95%	93.9%	93.3%
KPCA	96.9%	96.9%	96.9%	96.9%
KDA	96.9%	96.9%	96.9%	96.9%

The tables 2 and 3 are shown the results having applied the PCA/LDA/KPCA /KDA on DCT and FFT coefficients after taking the overall averages coefficients for all the training images, and the average of coefficients for each class.

Table 2: Accuracies on DCT Coefficients

DCT	Average of All Training Set		Average of Each Classes	
	DIST	KNN	DIST	KNN
PCA	94%	94%	88%	88%
LDA	92.9%	92.9%	95.9%	95.9%
KPCA	94%	94%	95.9%	95.9%
KDA	95.9%	95.9%	92.9%	92.9%

Table 3: Accuracies and on FFT Coefficients

FFT	Average of All training set		Average of each classes	
	DIST	KNN	DIST	KNN
PCA	41.8%	45%	60%	58%
LDA	88%	88%	87%	86.9%
KPCA	55%	55%	52.7%	52.7
KDA	55%	55%	62%	60%

Note from Table 1 and 2, that the performance of all features selection methods with recognition techniques (measure distance and KNN) are better, it is noticeable that Kernel methods is relatively better than linear methods. In table 3, we note that most of the testing results are less than the in Table 1 and 2, with the exception of the LDA have yielded a significant result with FFT coefficients.

Through the experiments, we are calculate the time of each method according to the training data as in tables 4 and 5.

Table 4: Time Complexity of Training Images Set in Average all Training Data

Average all Data(sec)	Direct	DCT	FFT
PCA	0.258	0.0177	0.0832
LDA	0.0642	0.0196	0.043
KPCA	0.029	0.012	0.0187
KDA	0.0337	0.0178	0.0246

Table 5: Time Complexity by Average per Classes Data

Averages Each Class(sec)	Direct	DCT	FFT
PCA	0.0253	0.0125	0.0152
LDA	0.0177	0.0113	0.0116
KPCA	0.0059	0.004	0.0047
KDA	0.0156	0.013	0.0108

We have observed that KPCA take less time than other methods, and the less time for KPCA method is 0.004 in with the experiment of average of per class with DCT coefficients. We noticed also the PCA take more time than other method and large time is 0.258 with the experiment of average of all images directly.

CONCLUSIONS

In this paper, we have proposed face recognition system based on linear PCA, LDA and non- linear KPCA, KDA methods to select an appropriate and reduction features .We had compared between these methods in terms of their ability to reach to the significant recognition performance. The distance measure and KNN methods are used in recognition phase .In their experiments, we notice that the linear and non-linear method achieved best results (Regardless of it be applied directly to the image data or indirectly after doing transfers to the image). Although the kernels methods is able to give the better performance when applied directly to the image data or after the transformation ,but it is not able in some cases to give the same better performance as with FFT coefficients. The experiment with less performance was with PCA +FFT+DIST with Average FFT training set, and the highest performance with KPCA and KDA in all the experiments in table 1 and 2.

REFERENCES

1. *I. Kim, J. H. Shim and J. Yang ,' Face detection', Standford university, [1998].*
2. *A. Abdallah, M. Abou El-Nasr and A. Lynn Abbott, "A New Face Detection Technique using 2D DCT and Self Organizing Feature Map" in Proc. of World Academy of Science, Engineering and Technology, Vol. 21, May [2007], pp. 15-19.*
3. *"Facial Recognition: Who's Tracking You in Public?". Consumer Reports. Retrieved 2016-04-05.*
4. *P. Wagner, 'Face Recognition with Python', July 18, [2012].*
5. *L. S. Balasuriya ,'Frontal View Human Face Detection and Recognition', B.Sc. thesis ,Department of Statistics and Computer Science, University of Colombo, Sri Lanka, May ,2000.*
6. *Z. Wang1, Q. Zhao, D. Chu, F. Zhao and L. J. Guibas5 ,'SELECT INFORMATIVE FEATURES FOR RECOGNITION', CA, USA ,[2010].*
7. *E. Hjelmas and B. K. Low, 'Face Detection: A Survey', Computer Vision and Image Understanding 83, 236–274, [2001].*
8. *O. H. Jensen ,'Implementing the Viola-JonesFace Detection Algorithm', IMM-M.Sc.: ISBN 87-643-0008-0 ,[2008].*
9. *P. F. Carrera, 'Face Recognition Algorithms', June 16,[2010].*

1. *10.S. A.Khayam ,'The Discrete Cosine Transform(DCT):Theory and Application', Department of Electrical & Computer Engineering, Michigan State University,March ,[2003].*

10. *E. F. Glynn, 'Fourier Analysis and Image Processing', Scientific programmer ,Bioinformatics ,STowers Institute for medical Research ,[2007].*
11. *Haykin S. ,Neural Networks: A comprehensive foundation, Prentice Hall, 0-13-273350-1,New Jersey, [1999].*
12. *A. R. dos Santos and A. Gonzaga, 'Automatic Clusters to Face Recognition' ,Department of Electrical Engineering, University of Sao Paulo ,Sao Carlos, Brazil.*
13. *I. K. Timotius, I. Setyawan, and A. A. Febrianto3,'Face Recognition between Two Person using Kernel Principal Component Analysis and Support Vector Machines', International Journal on Electrical Engineering and Informatics - Volume 2, Number 1, [2010].*
14. *Jayanthi T. and Dr. Aji S.,' Face Recognition Using KernelPrincipalComponent Analysis ',Advances in Vision Computing: An International Journal (AVC) Vol.1, No.1, March .[2014].*
15. *T.k. Kim, J. Kittler,'Locally Linear Discriminant Analysis for Multi-modally Distributed Classes For Face Recognition with a Single Model Image', IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.27, No.3, pp.318–327, [2005].*
16. *C. Li and B. Wang ,' Fisher Linear Discriminant Analysis', August ,31, [2014].*
17. *J. Yanga, Z. Jina, J. Yanga, D.Zhangc, and A. F. Frangib, 'Essence of kernel Fisher discriminant: KPCA plus LDA', The journal of Pattern Recognition ,[2004].*
18. *V. Roth and V. Steinhage, 'Nonlinear discriminant analysis using kernel function' ,In NIPS 12, pages :568 -574 ,[2000].*
19. *A. Singh, A. Yadav, A. Rana, 'K-means with Three different Distance Metrics', International Journal of Computer Applications, Volume 67, No.10, April, [2013].*
20. *O. Sutton,' Introduction to k Nearest Neighbour Classification and Condensed Nearest Neighbour Data Reduction', February, [2012].*

21. *L.Kozma,'k Nearest Neighbors algorithm (kNN)', Helsinki University of Technology, T-61.6020 Special Course in Computer and Information Science,[2008].*