ISSN: 2321-7782 (Online)
Volume 2, Issue 2, February 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Paper / Case Study
Available online at: www.ijarcsms.com

Experimental Assessment of LDA and KLDA for Face Recognition

Umesh Ashok Kamerikar[1]
Electronics & Telecommunication
SSPM's College of Engineering
Kankavli – India

Dr. M.S. Chavan[2]
Electronics & Telecommunication
Kolhapur Institute of Technology
Kolhapur – India

Abstract: Face recognition is a biometric technology with a wide range of applications such as access control, banking, information security, , database retrieval etc. This paper addresses the building of face recognition system by using Linear Discriminant Analysis (LDA) and Kernel LDA method. LDA is a statistical approach used for reducing the number of variables in face recognition. LDA is an enhancement to PCA. LDA and KLDA constructs a discriminant subspace that minimizes the scatter between images of same class and maximizes the scatter between different class images then classification is done by distance measure methods such as Euclidean distance. A number of trials were conducted to evaluate the performance of the face recognition system and compared the results.

Keywords: Covariance matrix, Face database, Face Recognition, Kernel Linear Discriminant Analysis (KLDA), Linear Discriminant Analysis (LDA).

I. INTRODUCTION

Face is one of the most important visual objects in our life which playing a major role in conveying identity and emotion and includes huge information. Face recognition is a vast research area in pattern recognition, computer vision and plays a important role in the applications of image analysis. Face recognition commonly includes feature abstraction, feature decrease and classification or recognition. Feature extraction is to find the most representative information of the faces, making them easily distinguishable from others. Face reduction is to not only extracts and compress the original features but also does not destroy the most important information of raw data. Recognition or classification is to choose the measure method which is used to classify the feature of images present in the database and test image such as Euclidean distance. Due to high dimensionality of face image it is difficult to use the original data as it is hence it is critical to choose the effectively distinguished features for extraction and reduction. LDA is one of effective feature extraction method based on face as a global feature. It decreases the dimension of image effectively and holds the primary information. In this paper face recognition system is described and it is followed by the LDA and KLDA algorithm using Euclidian distance classifier. Number of tests conducted using both algorithms and their characteristics are studied on the basis of evaluation parameters.

II. FACE RECOGNITION SYSTEM

Face recognition system consist of important steps viz. acquisition the test and train face data, face feature extracting and face recognition or classification.

In the face recognition system Acquisition and Processing of Face Data is first step. In this step face images is collected from database available or real time image using camera. For real time image many times processing of face database require. If it is not done it causes effect on the performance of systems as there are changes in the background, illumination condition, camera distance, lighting conditions, size and orientation of the head.

Mathematical representation of original image called biometric template. It is generated first and stored in the user train or test database. Using it formation of the basis (vector) of any recognition task is performed. These extracted features are used in recognition.

After feature extraction and selection images are classified according to features. Appearance-based face recognition algorithms use a wide variety of classification methods such as PCA, LDA. Similarity between faces from the same individual and different individuals will be considered for recognition.

III. Approach Taken for Recognition

The work presented here provides approach to the face recognition problem, describing face recognition system that can be used in application of Human computer interface. System contains main components as Feature Extraction, LDA and Kernel LDA algorithm and Euclidean Distance Classifier. To classify the image final face recognition system uses Euclidean Distance Classifier. The system developed is able to find and recognize the face of public database ORL (gray scale), Indian and Grimace (color) database.

IV. System Architecture

This section describes face recognition system architecture. The facial recognition system can be divided into four main stages: creating database, pre-processing data, feature extraction and classification. Fig.1 represents the basic blocks of Face recognition system.

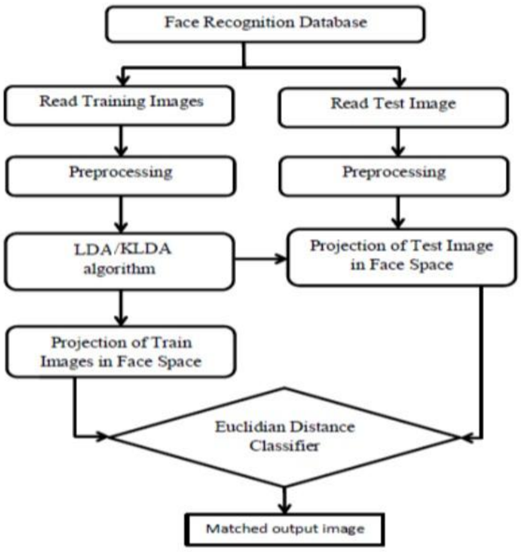

Fig 1: Face recognition System Architecture

V. Linear Discriminant Analysis (LDA)

PCA is based on the sample covariance which characterizes the scatter of the complete data set, without considering class separation. PCA might not provide good discrimination power with its projection axes selection. Perform dimensionality reduction while preserving as much of the class discriminatory information as possible LDA is the solution. Fisher Linear Discriminant (FLD) analysis, also called Linear Discriminant Analysis (LDA) finds the line that best separates the points. It performs face recognition by means of grouping images of the same class and separate images of different classes. [1,2]

The LDA method, which creates an optimal projection of the dataset, maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples. The within-class scatter matrix, represents variations in appearance of the same individual due to different lighting and expression changes, while the between-class scatter matrix, represents variations in appearance due to a difference in identity. In this way fisherfaces can project away some variation in lighting and facial expression while maintaining discriminability.

Solution to the singularity problem (Belhumeur, 1997) : Fisherface method

Fisherface method: Project the image set to a lower-dimensional space to avoid non singularity in the within-class scatter (S_W) Use PCA to achieve dimension reduction and apply the Fisher Linear Discriminant (FLD). Following are the steps to be required to be followed.[3]

1. Acquire the initial set of face images (the training set). Suppose $T_1, T_2, T_3, T_4, --------T_p$ (mxn) are (mxn) vectors and convert them from 2 dimension (mxn) to 1 dimension (mxn) xp)

2. Find the average face i.e. mean of Input Train Images

$$M = \frac{1}{p}\sum_{i=1}^{p} T_i$$

3. Find deviation of all train images from mean image.

$$A_i = T_i - M \quad (mnxp)$$

4. Compute covariance matrix

$C = \frac{1}{p}\sum_{n=1}^{p} \Phi_n \Phi_n^T = AA^T$ of size $(mnxmn)$

Matrix $A * A^T$ very large----not practical as size is large

$A * A^T$ and $A^T * A$ have same eigenvalues and their eigenvectors are related as $U_i = A * V_i$ hence find $L = A^T * A$ which is surrogate of covariance matrix

5. Compute the eigenvalues and eigenvectors of L and consider Eigen vectors whose eigenvalues are significant.

6. Compute the eigenvectors of C which are related to L eigenvectors as

$$U_i = A * V_i \; where \; i = 1,2,3 \dots\dots\dots\dots p \;\; of \; size(mnxN)$$

$N \leq p$ as highest values of eigenvalues and corresponding eigenvectors

7. Transfer train data to face space.

It is formed by taking the transpose of the vector and multiplies it with original data set

$$V\_PCA_i = (U^T x \, A_i) \quad where \quad i = 1,2,---p$$

Projection in to face pca as

$$\Omega_P = (V\_PCA^T x \, A_i)$$

8. Calculate the mean of each class (μ_i) and mean image of all classes (μ).

$$\boldsymbol{\mu_i} = \frac{1}{C_p}\sum_{j=1}^{C_p} x_j \quad where \;\; i = 1,2,---C_n, j = 1,2,--C_p$$

$$\boldsymbol{\mu} = \frac{1}{C_n}\sum_{j=1}^{C_n} \mu_j \qquad where \; j = 1,2,--C_n$$

9. Calculate within scatter matrix (s_w) & Between-class scatter matrix (s_B).

$$S_w = \sum_{i=1}^{C_n}\sum_{j=1}^{C_p}(x_j - \mu_i)(x_j - \mu_i)^T$$

$$S_b = \sum_{i=1}^{C_n}(\mu_i - \mu)(\mu_i - \mu)^T$$

10. Calculate eigenvector of J (J_eig_vec) from within scatter matrix (s_w) & Between-class scatter matrix (s_B)

11. Sorting of eigenvector of J

We have to sort out eigen vector depending upon their corresponding eigen values here this eigen values of J(w) are known as Fisher values and their corresponding eigenvector are known as Fisher vector. We neglect fishervector corresponding to small fishervalues

Sorted eigenvectors of J = V_Fisher

12. Project data in fisher space

Ω_p is converted into Ω_F by projecting onto a Fisher subspace, so that images of same class or person move closer together & images of different classes move further apart.

$$\Omega_F = (V\_Fisher^T x\ \Omega_p\ )$$

13. For testing Acquire the Test face images (the training set). Suppose T be the test image of size *(mxn)*, convert it from two dimensions (*mxn*) to one dimension *(mxn) x1)* and find deviation of Test image from mean image.

$$X = T - M\ (mnxp)$$

14. Transfer Test data to face space

$$\Omega_{NEW} = (V\_Fisher^T x V\_PCA^T\ x\ X)$$

15. Calculate minimum Euclidean distance

$$€_i = \min\ (\|\Omega_{NEW} - \Omega_i\|)$$

16. Find the index number of minimum Euclidian distance which is the best match for the Test image.

VI. Kernel Linear Discriminant Analysis (KLDA)

KERNEL is technique that takes into account the higher order statistics. Therein the input space is first mapped into feature space via nonlinear mapping and then principal components are determined in the feature space. They can compute the dot products of two feature vectors without even knowing what they are. Any algorithm which can be expressed solely in terms of dot products, i.e. without explicitly usage of the variables themselves, the kernel method enables us to construct nonlinear versions of it. Because of increase in dimensionality, the mapping is made implicit (and economical) by the use of kernel functions satisfying Mercer's theorem i.e. has non-negative eigenvalues.

$$k_{ij} = \{k(x_i, x_j)\} = \langle\Phi(x_i), \Phi(x_j)\rangle \text{in feature space i, j} = 1,2,3 - - - \text{N}$$

Kernel evaluations k_{ij} is mapping from input space to inner dot-products in the higher dimensional feature space.

Consider a Dataset $\mathbf{X}$ in high dimensional space $\Re^n$ $X_1, X_2, X_3, \ldots \ldots . X_n$ $\Re^n$, via the nonlinear mapping

Φ: $\Re^n \rightarrow F$ $\quad$ $X \rightarrow \Phi(x)$

X is mapped into a very high (possibly infinite) feature space F. we use nonlinear kernel function Φ(x) to project input data X into feature space F which is nonlinearly related to the input space and performs PCA in feature space F.

The kernel implemented is Gaussian Radial Basis function (RBF)

$$k(x\,,y) = exp\left(-\frac{1}{2\sigma^2}\|x-y\|^2\right)\ \ where\ \sigma = varience$$

In the kernel equation $(x\text{-}y)^2$ is the squared L_2-distance between two realizations. The kernel width parameter □ controls the flexibility of the kernel. A larger value of □ allows more "mixing" between elements of the realizations, whereas a smaller value of □ uses only a few significant realizations.

We can use the kernel idea for LDA [4,5] to find non-linear directions by first mapping the data non-linearly into some feature space F and computing Fisher's linear discriminant there, thus implicitly yielding a non- linear discriminant in input space. Consider the matrix which contains the non-linear mappings of all the training samples. To find the linear discriminant in F we need to maximize

$$j(w) = (w^T s_b^{\Phi} w)/(w^T s_w^{\Phi} w\ )$$

Again, this problem can be reduced to a eigenvalue problem of the same form as in LDA. Instead of mapping the data explicitly into the feature space, we seek a formulation of the algorithm which uses only the dotproducts of $\Phi(x_i).\Phi(x_j)$the images of training patterns in feature space. We are then able to compute these dotproducts efficiently without mapping explicitly to F. the cost function in this can be reduced to the form as

$$\mathrm{j}(\alpha) = (\alpha^T kQk\alpha)/(\alpha^T kRk\alpha)$$

Following are steps to implement KPCA algorithm

1. Acquire the initial set of face images (the training set). Suppose$T_1, T_2, T_3, T_4, -------- T_p$ (*mxn*) are (*mxn*) vectors and convert them from 2 dimension (*mxn*) to 1 dimension (*mxn) xp*)

2. Find the average face i.e. mean of Input Train Images

$$\mathrm{M} = \frac{1}{p}\sum_{i=1}^{p} T_i$$

3. Find deviation of all train images from mean image.

$$A_i = T_i - \mathrm{M}\quad (mnxp)$$

4. Construct Gaussian Radial Basis Function Kernel

$$k(x\,,y) = exp\left(-\frac{1}{2\sigma^2}\|x-y\|^2\right)\qquad (pxp)$$

5. Center the K matrix as

$$\widetilde{K} = k - 1_n k - k1_n + 1_n k1_n$$

where 1_n is (n x n) matrix with each value as 1/ n .

6. Compute the eigenvalues and eigenvectors of $\widetilde{K}$ and consider Eigen vectors whose eigenvalues are significant.

7. Compute the eigenvectors of C which are related to $\widetilde{K}$ eigenvectors as

$$U_i = A * V_i\ where\ i = 1,2,3 \dots\dots\dots p\ \ of\ size(mnxN)$$

$N \leq p$ as highest values of eigenvalues and corresponding eigenvectors

8. Perform 7 to 16 steps of LDA algorithm hereon.

VII. DATABASE USED FOR RECOGNITION

A. ORL /ATT Database

We have the proposed algorithm using human face images from Olivetti-Oracle Research Lab (ORL). The ORL dataset consists of 400 frontal faces: 10 tightly-cropped images of 40 individuals with variations in pose, illumination, facial expression (open/closed eyes, smiling/not smiling) and accessories (glasses/no glasses). The size of each image is 92x112 pixels, with 256 gray levels per pixel in Portable Gray Map (PGM) format. Figure shows some randomly selected images from the dataset for face recognition.[6]

Fig 2: Sample from ORL Database

B. Indian Database

This database contains images of 40 distinct subjects with 11 different poses for each individual. All the images have a bright homogeneous background and the subjects are in an upright, frontal position. There are variations in pose for the face looking front, looking left, looking right, looking up, looking up towards left, looking up towards right, looking down with variation in emotions. The files are in 24 bit RGB JPEG landscape format. The size of each image is 640x480 pixels. The images are organized in two main directories - males and females.[7]

Fig 3: Sample from Indian Database

C. Grimace Database

This database contains human face images in 24 bit RGB, JPEG format, 180x200 pixel portrait formats with plain background. There are 20 individuals each having 20 images with small head scale variation, very little Image lighting variation. Database contains images of male and female subjects, images of people of various racial origins major expression variation with glasses ad beards. Camera used is S-VHS camcorder. Lighting used is artificial, mixture of tungsten and fluorescent overhead.

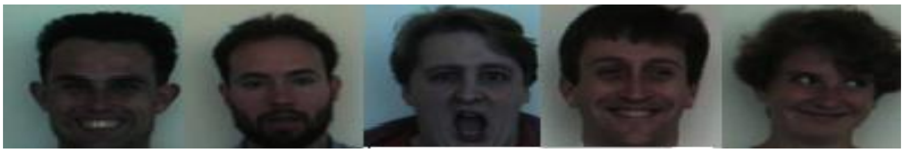

Fig 4: Sample from Grimace Database

The training set is consisted of 70 images (the set contains 10 persons and each person contains 7 images). On the other hand, the test set contains 70 images that are consisted of random choosing 10 images from every expression. The experiment is iterated 10 times.[8]

VIII. TOOL USED FOR EVALUATION

A. Correct Recognition accuracy

It is indication of accuracy of face recognition system. In this number of tests are conducted and out of which how many results are wrong i.e. erroneous found. Using it correct recognition rate is calculated using following formula

$$Correct\ Recognition\ accuracy = \left[1 - \frac{Error\ Count}{Total\ Test\ Image\ Count}\right] X\ 100$$

If it is 100% then that algorithm is accurate which ideal requirement is. But as train database is small correct recognition rate gets affected

B. Mean Square Error (MSE)

In statistics, the mean squared error (MSE) of an estimator is one of many ways to quantify the difference between values implied by an estimator and the true values of the quantity being estimated. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. MSE measures the average of the squares of the "errors." The error is the amount by which the value implied by the estimator differs from the quantity to be estimated. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate.

Let X and Y two arrays of size NxM respectively representing the X-channel frame of reference (the original copy) and Y-channel frame of the encoded/impaired copy then mean square error between the two signals is defined as

$$MSE = \frac{1}{NxM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} [X(i,j) - Y(i,j)]^2$$

The more Y is similar to X, the more MSE is small. Obviously, the greatest similarity is achieved when MSE equal to 0.[9]

C. Peak Signal to Noise Ratio (PSNR)

PSNR is most commonly used to measure the quality of reconstruction of lossy compression codecs (e.g., for image compression). The signal in this case is the original data, and the noise is the error introduced by compression. When comparing compression codecs, PSNR is an approximation to human perception of reconstruction quality. Although a higher PSNR generally indicates that the reconstruction is of higher quality, in some cases it may not. One has to be extremely careful with the range of validity of this metric; it is only conclusively valid when it is used to compare results from the same codec (or codec type) and same content.

PSNR is most easily defined via the mean squared error (MSE). It is given as Peak signal to noise ratio is given as,

$$PSNR = 10 \log_{10} \frac{L^2}{MSE}$$

L is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255. More generally, when samples are represented using linear PCM with B bits per sample, L will be 2^B-1. For color images with three RGB values per pixel, the definition of PSNR is the same except the MSE is the sum over all squared value differences divided by image size and by three. It is represented in decibels. A small mean square error results in a high signal to noise ratio, if MSE tends to zero, then PSNR tends to infinity.[10]

IX. Results Obtained using PCA and KPCA

In this case the performance of LDA and KLDA based feature extracted algorithm using Euclidian Distance classification has been examined. The performance is evaluated by examining the system's capability to distinguish between persons with different facial variations based on analyzing the facial expressions. The system has been tested over the ORL (gray scale), Indian and Grimace (color) images. In this testing, the system has examined its ability to distinguish between different personalities.

1: ORL database –Number of Train images per person varied from 1 to 10 and in every case 400 tests conducted which gave following recognition rate results

Table I: Recognition rate of LDA & KLDA

No. of Train Images per person	Image size (n x m) pixels	Accuracy Rate using LDA	Accuracy Rate using KLDA
1	92 x 112	20.00%	16.00%
2	92 x 112	31.25%	30.25%
3	92 x 112	47.00%	63.25%
4	92 x 112	44.00%	63.25%
5	92 x 112	56.25%	68.00%
6	92 x 112	65.25%	80.75%
7	92 x 112	77.50%	92.25%
8	92 x 112	85.25%	93.50%
9	92 x 112	92.25%	96.25%
10	92 x 112	100.00%	100.00%

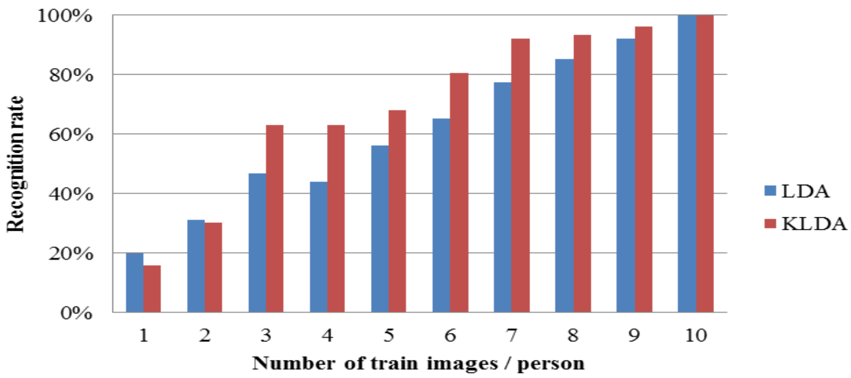

Fig. 5: Recognition rate of PCA & KPCA for ORL database

2. ORL database, Indian database and Grimace database –Training images 2/person For ORL 400 tests, Indian 200 and Grimace 288 tests conducted and following results received.

Table II: Recognition rate of LDA & KLDA

No. of Train Images per person	Image size (n x m) pixels	Accuracy Rate using LDA	Accuracy Rate using KLDA
ORL Database			
2	92 x 112	31.25%	30.25%

Indian Database			
2	640x480	15.00%	18.00%
Grimace Database			
2	180x200	8.33%	4.51%

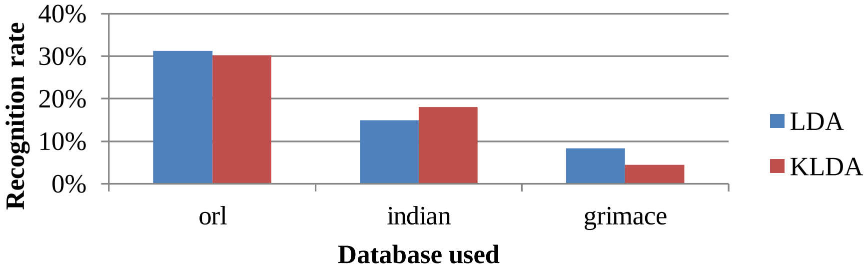

Fig 6: Recognition rate of LDA & KLDA

3. Average recognition rate for ORL Database with variation in Train database images/person from 1 to 10.

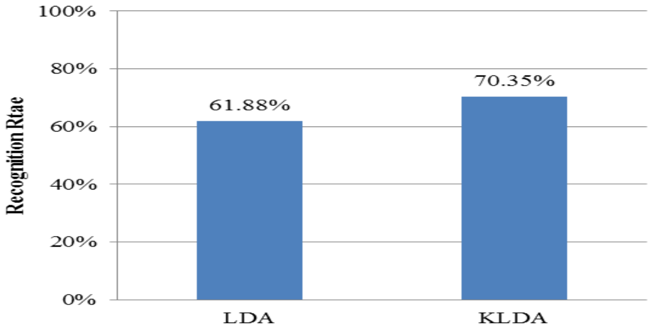

Fig 7: Average recognition rate of PCA & KPCA

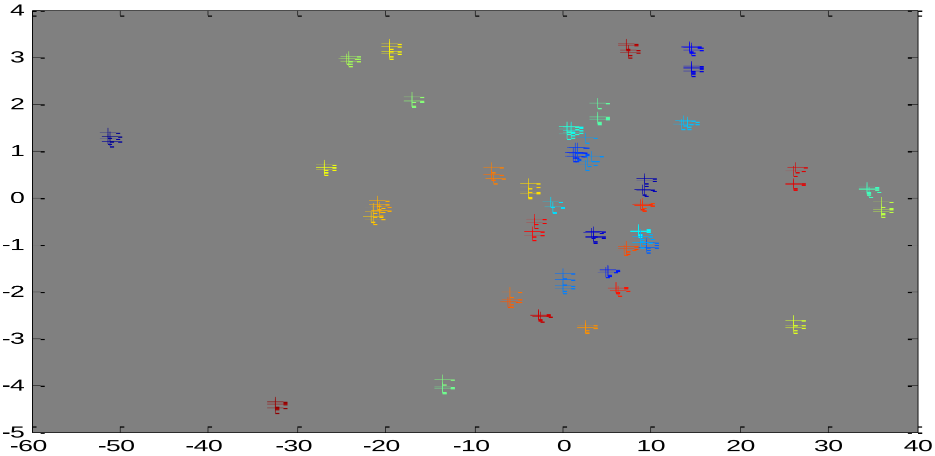

Fig 8: Projection of Train Images in Face space in LDA

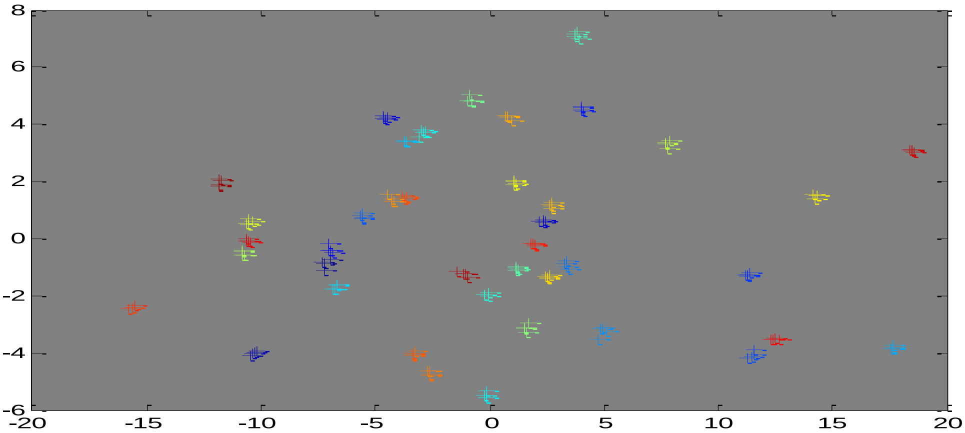

Fig 9: Projection of Train Images in Face space in KLDA

X. CONCLUSION

We have implemented facial recognition system using LDA and KLDA method. These two approaches have been compared for different database i.e. gray scale and color. From the result following are conclusions.

1. In between LDA and KLDA as no of Train images per person increases, recognition rate increases for ORL (gray scale) database but LDA works better way when less number of Train images per person chosen.

2. For any database whether gray scale or color for small number of Train images per person LDA outperforms than KLDA.

3. Projection of Train images for KLDA scatter within-classes is closer than LDA but also the scatter between-classes are farther than LDA.

4. Average recognition rate of KLDA performance is better than LDA.

5. When image is perfectly matched by any algorithm MSE is zero whereas PSNR infinity.

References

1. Fei Ye, Zhiping Shi , Zhongzhi Shi, A Comparative Study of PCA, LDA and Kernel LDA for Image Classification, 2009 International Symposium on Ubiquitous Virtual Reality.
2. Belhumeur, P.,Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 711–720
3. Jian Yang, Alejandro F. Frangi, Jing-yu Yang, David Zhang, Senior Member, IEEE, and Zhong Jin, KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 2, February 2005
4. Jian Huang, Pong C. Yuen, Member, IEEE, Wen-Sheng Chen, and Jian Huang Lai, Choosing Parameters of Kernel Subspace LDA for Recognition of Face Images Under Pose and Illumination Variations, IEEE Transactions on systems, man, and cybernetics—Part B: Cybernetics, Vol. 37, No. 4, August 2007
5. D.N. Zheng, J.X. Wang, Y.N. Zhao, and Z.H. Yang, An Iterative Algorithm for KLDA Classifier, World Academy of Science, Engineering and Technology, International Journal of Computer, Information Science and Engineering Vol:1 No:4, 2007
6. ORL /ATT database http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
7. Indian Face Database http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase /
8. Grimace database www.essex.ac.uk/mv/allfaces/grimace.html
9. Mean square error http://en.wikipedia.org/wiki/Mean_squared_error
10. Peak signal to noise ratio http://en.wikipedia.org/wiki/PSNR

AUTHOR(S) PROFILE

Umesh Kamerikar completed his Bachelor of Engineering in Electronics Engineering from Rajarambapu Institute of Technology in 1999 and presently persuing M.E. from same institute. He is having teaching experience of 12 years with specialization in image processing, Control system, Microprocessor, microcontroller and communication.