Combination of Nonlinear Dimensionality Reduction Techniques for Face Recognition System

S.K.Sandhu (M.E. (ECE))*, Sumit Budhiraja (Asst. Professor)**

Electronics and Communication Engineering Department,
UIET, Panjab University,
Chandigarh

ABSTRACT
A Face Recognition System is used to automatically identify or verify a person from digital image. Since capturing of face image is not very difficult process and does not require too much cooperation of the subject, it keeps the interest of researchers alive. In this paper, combination of linear and combination of nonlinear dimensionality reduction techniques are implemented separately for face recognition system. The linear methods used are PCA, LDA and LPP and nonlinear methods used are LLE and ISOMAP. Recognition Rate is analyzed for comparing both systems. The results obtained on ORL database show that system with combination of nonlinear techniques performs better than linear counterpart.

Keywords: **ISOMAP, LDA, LLE, LPP, PCA**

1. INTRODUCTION

Face Recognition System is the one of the biometric system which is used for security purposes all over the world. It is used to identify or verify a person from a digital image. Biometric systems are preferred from security point of view since they are dependent on what a person has rather than what he can remember. Since capturing an image is the easy process and does not require too much cooperation of the subject, it is topic of wide interest of researchers from almost a decade.

With fast increase in quantity and complexity of data, it becomes difficult rather impossible to directly deal with raw data. Also, irrelevant features reduce efficiency of the system. Hence, dimensionality reduction is important prior to identification. It basically extracts intrinsic structure of the data and discards the redundant data. Appearance based methods treat images as two dimensional intensity matrices and use statistical properties to analyse an image. E.g. the image of size $m \times n$ pixels becomes matrix of mn size. As number if images in training set increases, size of matrix increases,. This curse of dimensionality is reduced by dimensionality reduction techniques. Appearance based methods can be classified as linear methods and nonlinear methods. Linear methods explicitly transform data from high dimensional subspace into low dimensional subspace by linear mapping. However, the general problems faced in real time face recognition system are pose variations, illumination variations, differing environmental conditions, aging effects etc. All these variations tend to make face database nonlinear in nature. Linear methods fail to reveal the intrinsic structure of nonlinear data and also corresponding reconstruction error tends to rise. In nonlinear techniques, explicit projections are not done. Instead faithful low dimensional data matrix is obtained directly from high dimensional data matrix. Nonlinear methods have ability to deal with complex nonlinear data and so features retained help to increase the efficiency of the system.

In this paper, combination of linear methods i.e. Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Locality Preserving Projections (LPP) is used to implement linear system [1]. Combination of nonlinear methods i.e. Isometric Mapping (ISOMAP) and Locally Linear Embedding (LLE) is used to implement nonlinear system. Comparison between linear and nonlinear system is done based on recognition rate.

2. LINEAR DIMENSIONALITY REDUCTION TECHNIQUES

Linear Dimensionality Reduction Techniques used in this paper are explained in depth in following sections.

3.1. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical dimensionality-reduction method, which produces the linear least-squares subspace of a training set [2]. Given a y-dimensional vector representation of each face in a training set of M images, the PCA finds a z- dimensional subspace whose basis vectors correspond to the maximum variance directions in the original image space. New z-dimensional subspace is normally lower dimensional space than the original space (z<<y). Given a total of M images with (N_x x N_y) pixels, we convert them into training set Γ = [Γ1 Γ2 … ΓM] with lexicographic ordering the pixel elements. Difference matrix is the training data with their mean removed [3]. Covariance matrix is computed step-by-step from the difference matrix as given by following equations:-

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

$$\varnothing_i = \Gamma_i - \psi \quad (2)$$

$$A = [\varnothing_1\ \varnothing_2\ \ldots\ldots \varnothing_M] \quad (3)$$

$$C = A.A_T = \frac{1}{M}\sum_{i=1}^{N}\varnothing i\ \varnothing_T \quad (4)$$

where Ψ is mean of whole data in vector form, Ø is a mean subtracted image, and A is difference matrix. C is covariance matrix lying in a very high dimension which is (N_x x N_y)* (N_x x N_y). Solution to this problem is provided for, by using a covariance matrix L with small dimension which is (M x M). Eigenvectors, v computed from covariance L is multiplied with A to yield another variable υ which is able to represent the actual eigenvectors of covariance C.

$$L = A^T A = \frac{1}{M}\sum_{i=1}^{M}\Gamma_i.\Gamma_i^T \quad (5)$$

$$\upsilon = A.\upsilon \quad (6)$$

Weight sets corresponding to respective subjects in training set are obtained using projection basis defined by following equations:-

$$w_k = \upsilon_k^T.\varnothing = \upsilon_k^T.(\Gamma - \psi) \quad (7)$$

$$\Omega = [w_1\ w_2\ \ldots\ldots w_M] \quad (8)$$

where w_k is weight and Ω is weight set. Not all eigenvectors are needed, thus selecting M eigenvectors obtain projection basis. For single test image identification, mean subtraction is done and projected on υ_k to obtain its weight.

3.2 Linear Discriminant Analysis (LDA)

The Linear Discriminant Analysis (LDA) is also one of the most popular linear projection techniques. It finds the set of the most Discriminant projection vectors which map high-dimensional samples onto a low-dimensional space. Belhumeur et al. [4] firstly presented projection method based on the Fisher's Linear Discriminant (FLD) in 1997. For all samples of all classes, the between-class scatter matrix S_B and the within-class scatter matrix S_W are defined. The goal is to maximize S_B while minimizing S_W, in other words, maximize the ratio det| S_B | / det| S_W |. This ratio is maximized when the column vectors of the projection matrix are the eigenvectors of (S_W^{-1}· S_B).

Within – class and between-class scatter matrices are computed as follows:

$$S_W = \frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{N_k}[(X_k^i - x_k)(X_k^i - x_k)^T] \quad (9)$$

$$S_B = \frac{1}{M}\sum_{k-1}^{M}[N_k(x_k - x)(x_k - x)^T] \quad (10)$$

Here S_W is the within-class scatter matrix, S_B is the between class scatter matrix, X^i_k is image i of class k (k=1, 2, … M). x_k is the mean vector of class k, x is overall mean vector, N_k is the sample size of class k.

Various measures are available for quantifying the discriminatory power, Fisher criterion is the common one:

$$J(W) = \frac{Wt.Sb.W}{Wt.Sw.W} \quad (11)$$

Here **W** is the optimal projection matrix, which is obtained via solving the generalised eigenvalues problems:

$$S_B.W = \lambda\, S_W W \quad (12)$$

Mapping the training sample x_i into LDA subspace is as follows:

$$y_i = W^T.x_i \quad (13)$$

Here low dimension vector y_i is the LDA feature of the sample x_i. LDA has powerful discriminatory power. LDA produces well separated classes in low dimensional subspace even under severe variation in lighting and facial expression. However, It takes each image as point in high dimensional space where even little amount of geometric transform degrades recognition performance severely. LDA also suffers from small sample size problem which exists in high dimensional pattern recognition task where number of available sample is smaller than dimensionality of the samples.

3.3 Locality Preserving Projections (LPP)

Locality Preserving Projections are linear projective maps that arise by solving a variational problem that optimally preserves the neighborhood structure of the data set. LPP can be seen as an alternative to Principal Component Analysis (PCA). When the high dimensional data lies on a low dimensional manifold embedded in the ambient space, the Locality Preserving Projections are obtained by finding the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the manifold. As a result, LPP shares many of the data representation properties of nonlinear techniques such as Laplacian Eigenmaps or Locally Linear Embedding. Yet LPP is linear and more crucially is defined everywhere in ambient space rather than just on the training data points [5].

The algorithm contains the following steps:

1. Constructing the adjacency graph: Let G denote a graph with m nodes. An edge is put between nodes i and j if $\mathbf{x}_i$ and $\mathbf{x}_j$ are “close”. There are two variations:

(a) ε-neighborhoods [Parameter ε □ R]: Nodes i and j are connected by an edge if $||x_i - x_j||^2 < \varepsilon$ where the norm is the usual Euclidean norm in R^n.

(b) k nearest neighbors [Parameter k □ N]: Nodes i and j are connected by an edge if i is among k nearest neighbors of j or j is among k nearest neighbors of i.

2. Choosing the weights: Here, as well, there are two variations for weighting the edges. W is a sparse symmetric m × m matrix with W_{ij} having the weight of the edge joining vertices i and j, and 0 if there is no such edge.

(a) Heat kernel [Parameter t □ R]: If nodes i and j are connected, put

$$W_{ij} = e^{\frac{||x_i - x_j||^2}{t}} \quad (14)$$

(b) Simple-minded [No parameter]: $W_{ij} = 1$ if and only if vertices i and j are connected by an edge.

3. Eigen maps: Compute the eigenvectors and eigenvalues for the generalized eigenvector problem:

$$X.L.X^T a = \lambda.X.D.X^T.a \quad (15)$$

where D is a diagonal matrix whose entries are column (or row, since W is symmetric) sums of W, $D_{ii} = \sum j\, W_{ji}$. $L = D - W$ is the Laplacian matrix. The ith column of matrix X is $\mathbf{x}_i$. Let the column vectors a_0;__ _ _;$\mathbf{a}_{l-1}$ be the solutions of equation (15), ordered according to their eigenvalues, $\lambda_0 <$ _ _ _ $< \lambda_{l-1.}$ Thus, the embedding is as follows:

$$x_i \rightarrow y_i = A^T.x_i\,, \quad A = (a_0;\ a_1; \ldots\ldots a_{l-1}) \quad (16)$$

where $\mathbf{y}_i$ is a l-dimensional vector, and A is a n × l matrix. $\mathbf{y}_i$ represents Laplacian faces. LPP is good method to preserve local unique features from an image.

3. NONLINEAR DIMENSIONALITY REDUCTION TECHNIQUES

Nonlinear Dimensionality Reduction Techniques used in this paper are explained in subsequent sections:

3.1 Isometric Mapping (ISOMAP)

Classical linear scaling has proven to be successful in many applications, but it suffers from the fact that it mainly aims to retain pair wise Euclidean distances, and does not take into account the distribution of the neighboring data points. If the high-dimensional data lies on or near a curved manifold, classical scaling might consider two data points as near points, whereas their distance over the manifold is much larger than the typical inter-point distance. ISOMAP is a technique that resolves this problem by attempting to preserve pair wise geodesic (or curvilinear) distances between data points [6].

Geodesic distance is the distance between two points measured over the manifold. The great-circle distance or geodesic distance is the shortest distance between any two points on the surface of a sphere measured along a path on the surface of the sphere. For data lying on a nonlinear manifold, the "true distance" between two data points is the geodesic distance on the manifold, i.e., the distance along the surface of the manifold, rather than the straight-line Euclidean distance [7]. The main purpose of ISOMAP is to find the intrinsic geometry of the data, as captured in the geodesic manifold distances between all pairs of data points. The approximation of geodesic distance is divided into two cases. In case of neighboring points, Euclidean distance in the input space provides a good approximation to geodesic distance. In case of faraway points, geodesic distance can be approximated by adding up a sequence of "short hops" between neighboring points. ISOMAP shares some advantages with PCA, LDA, and MDS, such as computational efficiency and asymptotic convergence guarantees, but with more flexibility to learn a broad class of nonlinear manifolds. The ISOMAP algorithm takes as input the distances d (x_i, x_j) between all pairs x_i and x_j from N data points in the high-dimensional input space R^q. The algorithm outputs coordinate vectors y_i in a d-dimensional Euclidean space R^d that best represent the intrinsic geometry of the data.

The detailed steps of ISOMAP are listed as follows [8]:

1. Construct neighborhood graph: Define the graph G over all data points by connecting points x_i and x_j and if they are closer than a certain distance ε, or if x_i is one of the -nearest neighbors of x_j . Set edge lengths equal to d (x_i, x_j).

2. Compute shortest paths: Initialize $d_G(x_i,x_j) = d(x_i,x_j)$ if x_i and x_j are linked by an edge; $d_G(x_i,x_j) = +\infty$ otherwise. Then, for each value of k=1,2,....N, in turn, replace all entries $d_G(x_i,x_j)$ by $\min\{d_G(x_i,x_j), d_G(x_i,x_k) + d_G(x_j,x_k)\}$. The matrix of final values $D_G = \{d_G(x_i,x_j)\}$ will contain the shortest path distances between all pairs of points in (this procedure is known as Floyd's algorithm). We can also use Dijkstra's shortest path algorithm.

3. Construct d-dimensional embedding: Let λp be the pth Eigen value (in decreasing order) of the matrix $\tau(D_G)$ (The operator τ is defined by $\tau(D) = -HSH / 2$, where S is the matrix of squared distances $\{S_{ij} = D_{ij}^2\}$, and H is the "centring matrix" $\{H_{ij} = \sigma_{ij} - 1/N\}$, σ_{ij} is the Kronecker delta function , and v_p^i be the ith component of the pth eigenvector. Then set the pth component of the d-dimensional coordinate vector y_i equal to $\sqrt{\lambda p}\, v_p^i$.

3.2 Locally Linear Embedding (LLE)

Locally Linear Embedding (LLE), first proposed by Saul and Roweis in 2000 [9], is also a nonlinear manifold learning method. It is a technique that is similar to ISOMAP in that it also constructs a graph representation of the data points. However, in contrast to ISOMAP, it attempts to preserve solely local properties of the data. As a result, LLE is less sensitive to short-circuiting than ISOMAP, because only a small number of local properties are affected if short-circuiting occurs. Furthermore, the preservation of local properties allows for successful embedding of non-convex manifolds. In LLE, the local properties of the data manifold are constructed by writing the high-dimensional data points as a linear combination of their nearest neighbors. In the low-dimensional representation of the data, LLE attempts to retain the reconstruction weights in the linear combinations as good as possible.

LLE describes the local properties of the manifold around a data point xi by writing the data point as a linear combination w_i (the so-called reconstruction weights) of its k nearest neighbors x_{ij}. Hence, LLE fits a hyper-plane through the data point x_i and its nearest neighbors, thereby assuming that the manifold is locally linear. The local linearity assumption implies that the reconstruction weights w_i of the data points x_i are invariant to translation, rotation, and rescaling. Because of the invariance to these transformations, any linear mapping of the hyper-plane to a space of lower dimensionality preserves the reconstruction weights in the space of lower dimensionality. In other words, if the low-dimensional data representation preserves the local geometry of the manifold, the reconstruction weights w_i that reconstruct data point x_i from its neighbors in the high-dimensional data representation also reconstruct data point y_i from its neighbors in the low-dimensional data representation.

Suppose that the sample data set is $X= [x_1,x_2 ,...... ,x_i ,....x_n]$, where $x_i \in R^{d\times 1}$, and $X \in R^{d\times N}$.

The output value Y can be obtained through LLE, where Y= [y, y_2 ...yi....y_n] where y $\in R^l$, $Y \in R^{l \times N}$ and l << d , and l is the embedding dimension of LLE. The detailed steps of LLE algorithm are as follows [10]:

1. Search for k nearest neighbors of each data point in high dimensional space. Distance formula can be expressed as

$$d_{ij} = \left[\sum_{k=1}^{K} |x_{ik} - x_{jk}|^q\right]^{\frac{1}{q}} \quad (17)$$

For Euclidean distance, q=2, k is total number of the neighbor points.

2. Calculate the local reconstruction weight matrix of the sample point. Reconstruction errors are measured by the function

$$\min\emptyset(W) = \sum_{i=1}^{N} \left\| x_i - \sum_{j=1}^{k} w_{ij}.x_{ij} \right\|^2 \quad (18)$$

Where w_{ij} must satisfy the equation $\sum_{j=1}^{k} wij = 1$, w_i=[w_{i1}, w_{i2} ...wij] and $w_i \in R^{N \times l}$. If x_{ij} does not belong to the neighbors of x_i, w_{ij} = 0. Thus weight matrix can be obtained, represented by W= [w_1, w_2... w_n] , where $W \in R^{N \times N}$. Compute the covariance matrix of the neighbor points of x_i, we have

$$C_{jm} = (x_i - x_{ij})(x_i - x_{im})^T \quad (19)$$

Therefore, we can get w_{ij} by solving linear equation

$$\sum_{m=1}^{k} \sum_{j=1}^{k} C_{jm}.W = 1 \quad (20)$$

3. Map all sample points to the low-dimensional space. The process should meet

$$\min\emptyset(Y) = \sum_{i=1}^{N} \left\| y_i - \sum_{j=1}^{N} w^{ij}.y^{ij} \right\|^2 \quad (21)$$

Where $\emptyset$ (Y) is the loss function. Considering the constraint condition

$$\sum_{i=1}^{N} y_i = 0\,; \quad (22)$$

$$\frac{1}{N}\sum_{i=1}^{N} y.y_i^T = 1 \quad (23)$$

y = arg min$\emptyset$ (Y) can be solved, where I is an l ×l unit matrix.

4. IMPLEMENTATION OF LINEAR AND NONLINEAR SYSTEMS

In Facial Recognition system, three folders are made. One has training data and other has test data. Training data includes number of images of persons in different conditions. Test data contains query image that has to be authenticated from the available database. Test data includes images of same persons as in training data but with different poses or expressions. Third folder has images of persons treated as impostors. This folder contains images of persons not included in training data set. As discussed above, Dimension reduction is important step in face recognition system to reduce redundancy of data. In this paper, combination of linear techniques and combination of nonlinear techniques are used to implement linear and nonlinear system respectively. The Flow Diagram followed for implementation of combination of linear methods and combination of nonlinear methods is shown Fig.1 and Fig.2 respectively.

PCA is a global approach that uses Eigen vectors and Eigen values for representing face images. Those Eigen values are considered as set of features which together characterize the variation between face images, so it preserves holistic features from face image. LDA has high discriminatory power. It produces well separated classes in low dimensional subspace. So it is good to extract discriminatory features between classes. LPP is a linear projective map that arises by solving variational problem that optimally preserves the neighborhood structure of the data set. It extracts unique features from dataset. It also preserves locality so it is good for quick retrieval in high dimensional space.

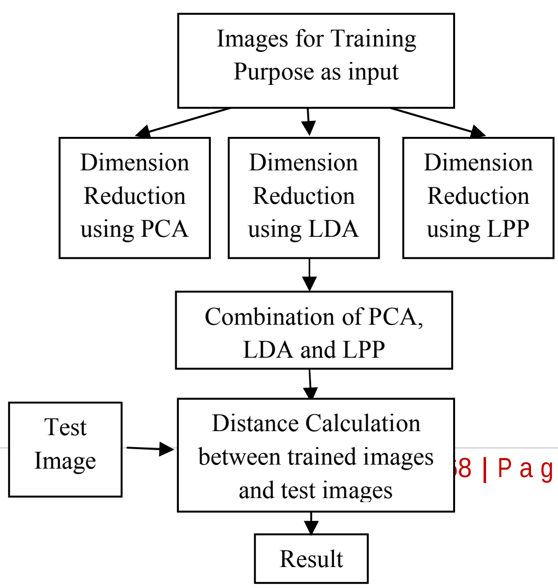

Figure1. Flow Diagram for combination of Linear Dimensionality Reduction Techniques

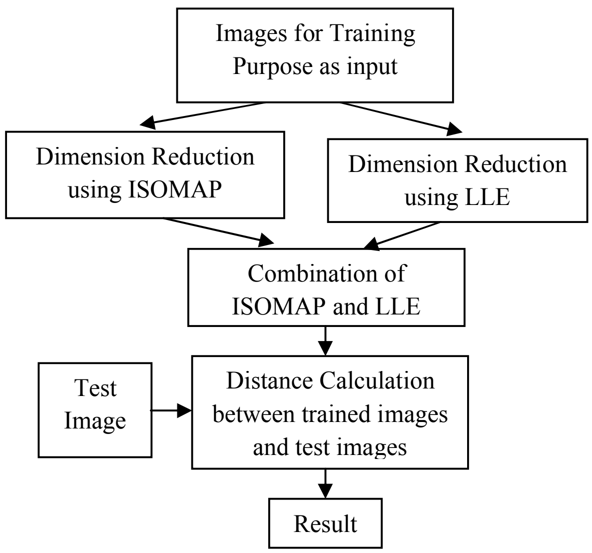

Figure2. Flow Diagram for combination of Nonlinear Dimensionality Reduction Techniques

Isomap is full spectral method that attempts to preserve pair wise geodesic distance instead of Euclidean distance between data points. So it extracts good global features. LLE is a sparse spectral technique that is similar to ISOMAP in the sense that it also constructs graph representation of data points. But in contrast to Isomap, it attempts to solely preserve local properties of data by writing high dimensional data points as linear combination of nearest neighbors. So it preserves local features well.

Projections obtained from PCA, LDA and LPP are combined and projected on one subspace for linear system. Projections obtained from ISOMAP and LLE are combined and projected on subspace for nonlinear system. Euclidean distance classifier is used to calculate distance between characteristics of test image (or query image) and training data set. The training set feature vector with least distance gives the best match image with the test image .If the minimum distance obtained is less than predefined threshold, person is authenticated otherwise person is treated as impostor.

5. RESULTS

In this paper, Recognition rate of linear and nonlinear system is calculated for comparison of performance of both systems. Recognition rate signifies accuracy of the face recognition system. It basically tells how efficiently our system distinguishes between persons in training system and impostors, i.e. how accurately it accepts persons with true identity and rejects other. Confusion matrix is deduced for each system as shown in Fig.3. True Positive (TP) indicates correct acceptance of authenticated person. True Negative (TN) indicates correct rejection of impostor or unauthenticated person. False Positive (FP) indicates rejection of authenticated person and false negative (FN) indicates acceptance of unauthenticated person or impostor. P indicates total number of authenticated persons and N indicates total number of unauthenticated persons.

$$TPR = \frac{TP}{P} = \frac{TP}{TP+FN} \tag{24}$$

$$FPR = \frac{FP}{N} = \frac{FP}{FP+TN} \tag{25}$$

$$\text{Recognition Rate} = \frac{TP+TN}{P+N} \tag{26}$$

The experiments are conducted on ORL face database [11]. Fig.4 shows the data set of all people in ORL database with different facial expressions. The ORL database image size is 92×112 which is small. For training of data, 7 images per person are used and randomly 2 images are used for testing purposes. MATLAB commands were used for programs with Intel(R) Core(TM) Duo CPU 2.20 GHz and 3GB RAM. Five different cases are considered for analysing behaviour of system with different number of images Table1 shows recognition rate obtained for system implemented using combination of linear dimensionality reduction techniques under different cases. Table 2 shows recognition rate obtained for system implemented using combination of nonlinear dimensionality reduction techniques under different cases.

From Table 1 and Table 2, it can be interpreted that for both systems, True Positive Rate and False Positive Rate increases with increase in number of training images. But there is overall increase in performance of both linear and nonlinear systems with respect to Recognition Rate.

True positive (TP)	False negative (FN)
False positive (FP)	True negative (TN)

Figure3 Confusion Matrix

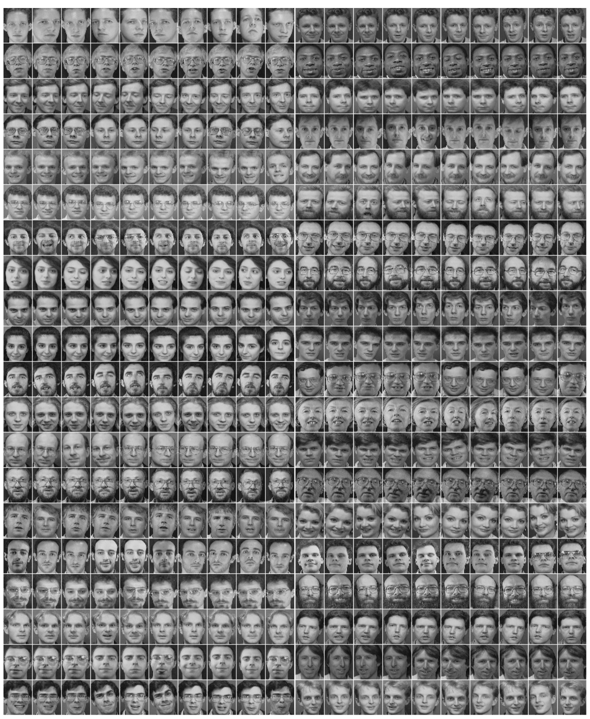

Figure4 ORL Database.

Table 1 Recognition rate of ORL data using combination of linear dimensionality reduction techniques

No. of training Images	No. of test images	No. of Impostors images	TPR	FPR	Total Recognition Rate
70	20	24	95%	14.28%	90.9%
105	30	32	93.3%	9.3%	91.93%
140	40	48	92.5%	8.3%	92.04%
170	50	64	92%	7.8%	92.11%
210	60	80	93.3%	7.5%	92.8%

Table 2 Recognition rate of ORL data using combination of nonlinear dimensionality reduction techniques

No. of training Images	No. of test images	No. of Impostors images	TPR	FPR	Total Recognition Rate
70	20	24	90%	4.16%	93.18%
105	30	32	93.3%	6.25%	93.54%
140	40	48	95%	6.8%	94.31%
170	50	64	96%	4.6%	95.61%
210	60	80	96.6%	5%	95.71%

Fig.5 shows graphical representation of comparison of results for both linear and nonlinear systems. It can be clearly seen from Fig.5 that performance of nonlinear systems is better than combination of linear system. Although Recognition Rate increases with increase in Number of images in training set, this increase is more for Combination of Nonlinear system implemented as compared with linear counterpart.

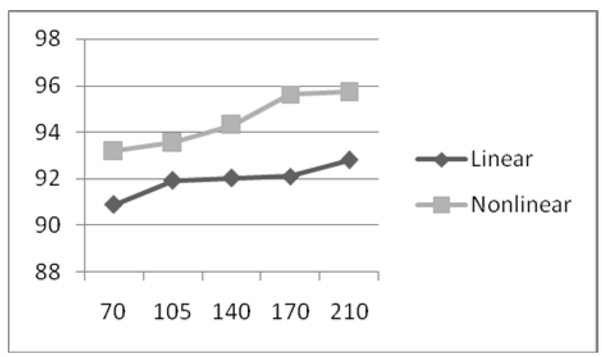

Figure5 Comparison of performance of linear and nonlinear systems with varying number of training image based on Recognition rate using ORL data

6. CONCLUSION

In this paper, Face Recognition system using combination of linear dimensionality reduction techniques and combination of nonlinear dimensionality reduction techniques is implemented for linear and nonlinear system respectively. Linear techniques used are PCA, LDA and LPP. Nonlinear techniques used are ISOMAP and LLE. Recognition rate of both systems is analysed with varying the number of images in training set. Comparison of recognition rates with varying number of images in training set for both systems individually and with each other is done. It is seen that recognition rate improves with increase in number of images in training set. Also, overall recognition rate is better for nonlinear system as compared with linear counterpart.

REFERENCES

[1] Shylaja S S , K N Balasubramanya Murthy, S Natarajan Pritha D N, L Savitha "Face Recognition Using Trichotomic Combination Of SVD, DF-LDA And LPP" The 2nd International Conference on Computer and Automation Engineering (ICCAE, 2010), Vol.1, pp: 246 – 250, Feb.2010.

[2] Hyunjong Cho, Seungbin Moon, "Comparison of PCA and LDA based face recognition algorithms under illumination variations", International Joint Conference ICCAS-SICE, pp: 4025 – 4030, Aug 2009.

[3] Lih-Heng Chan, Sh-Hussain Salleh, Chee-Ming Ting, A. K. Ariff,"Face Identification and Verification Using PCA and LDA", International Symposium on Information Technology, 2008(IT Sim 2008), Vol. 2, pp: 1 – 6,) Aug. 2008.

[4] P.N. Belhumeur, J.P. Hespanha, and D.J. Kiregman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 19, Issue 7, pp. 711- 720, 1997.

[5] Xiaofei He, Partha Niyogi, "Locality Preserving Projections", Advances in Neural Information Processing Systems (NIPS), pp:1-8, 2003.

[6] J.B. Tenenbaum, V. de Silva and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction", Science, Vol. 290, Issue 5500, pp. 2319–2323, 2000.

[7] Dianting Liu; Zongying Ou; Guoqiang Wang; Shungang Hua; Tieming Su; "Face Recognition Using Hierarchical Isomap", IEEE Workshop on Automatic Identification Advanced Technologies, pp: 103 – 106, June 2007.

[8] Xin Geng, De-Chuan, Zhan, Zhi-Hua Zhou, "Supervised Nonlinear Dimensionality Reduction for Visualization and Classification", Systems, IEEE Transactions on Man, and Cybernetics, Part B: Cybernetics, Vol. 35, Issue 6, pp. 1098 – 1107, Dec. 2005.

[9] S.T. Roweis and L.K. Saul, "Nonlinear dimensionality reduction by Locally Linear Embedding" Science, Vol. 290, Issue 5500, pp.2323–2326, 2000.

[10] Junying Gan; Pan Shao; Yibin Yu, "A research for face recognition based on locally linear embedding algorithm", Chinese Conference on Pattern Recognition (CCPR'2010), pp: 1 – 4, Oct. 2010.

[11] R. F. andez and E. Viennet, "Face identification using support vector machines", Proceedings of the 7th European Symposium on Artificial Neural Networks (ESANN), pp. 1-8, April 1999.