International Journal of Computer Theory and Engineering, Vol. 2, No. 4, August, 2010
1793-8201

Face Recognition Using Wavelet Based Kernel Locally Discriminating Projection

Venkatrama Phani Kumar S[1], KVK Kishore[2] and K Hemantha Kumar[3]

***Abstract*—Locality Preserving Projection(LPP) aims to preserve the local structure of the image space, while Principal Component Analysis(PCA) aims to preserve the global structure of the image space; LPP is linear, while Isomap, LLE, and Laplacian Eigenmap are nonlinear methods, so they yield maps that are defined only on the training data points and how to evaluate the maps on novel test data points remains unclear. Locally Discriminating Projection (LDP) is the extension of LPP, which seeks to preserve the intrinsic geometry structure by learning a locality preserving submanifold. LDP is a new subspace feature extraction method and supervised because it considers both class and label information. LDP performs much better than the other feature extraction methods such as PCA and Laplacian faces. In this paper an extension to LDP called Wavelet based Kernel Locally Discrimination Projection (-WKLDP) is proposed to extract non linear features of subband face images for classification, where as LDP considers linear features only. In the proposed method first by using wavelets the subband face images are constructed, then on subband face images kernel locally discriminating projection (KLDP) is applied. The experimental results on the ORL face database suggest that W-KLDP gives lower time complexity and have high recognition rates than other existing methods.**

***Index Terms*—Dimensionality Reduction, Locality Preserving Projection, Locally Discriminating Projection, Discrete Wavelet Transform, Wavelet based Kernel Locally Discrimination Projection.**

I. INTRODUCTION

Recently, due to large applications, there has been much interest in automatic recognition of faces. Face recognition can be defined as the identification of individuals from images of their faces by using a stored database of faces labeled with people's identities. Many face recognition methods have been developed over the past few decades. In these approaches, a two dimensional image of size w X h pixels is represented by a vector in a wh-dimensional space. Therefore, each facial image corresponds to a point in this space. This space is called the sample space or the image space, and its dimension typically is very high. A common way to attempt to resolve this problem is to use dimensionality reduction techniques.

Locality Preserving Projections (LPP) [5] is a new linear dimensionality reduction algorithm. Compared to other dimensionality reduction techniques, it has distinct advantages in two aspects: LPP preserves the local structure of the image space. LPP is an unsupervised feature extraction because it considers only class information. By considering label information along with class information a new feature algorithm is proposed called Locally Discriminating Projection. It is the extension of LPP and it is a supervised feature extraction algorithm. LDP is linear and considers only linear features. A new feature extraction algorithm called Kernel Locally Discriminating Projection proposed to consider nonlinear features.

Wavelets have been successfully used in image processing. Its ability to capture localized time-frequency information of images motivates its use for feature extraction. Numerous techniques have been developed to make the imaging information more easily accessible and to perform analysis automatically. Wavelet transforms [4] have proven promising useful for signal and image processing. Wavelets are applied to various aspects of imaging informatics, including image compression. Image compression is a major application area for wavelets. Because the original image can be represented by a linear combination of the wavelet basis functions, compression can be performed on the wavelet coefficients too.

The features of Wavelet transforms are given below:

Multiresolution: Wavelet transform analyzes the image at different scales or resolutions.

Locality: Wavelet transform decomposes the image into subbands that are localized in both space and frequency domains.

Sparsity: A wavelet coefficient is large if the singularities are present in the support of a wavelet basis function. The magnitudes of coefficients tend to decay exponentially across scale. Most energy concentrates on large coefficients.

Decorrelation: Wavelet coefficients of images tend to be approximately decorrelated because of the orthonormal property of wavelet basis functions.

These properties make the wavelet domain of natural image more propitious to feature extraction for face recognition, compared with the direct spatial-domain.

DB4 is an asymmetrical, orthogonal and biorthogonal wavelet transform. DB4 is used in this proposed method for image compression. The Scaling and Wavelet functions of Wavelet Daubechies 4 are given in figure 1. Decomposition of low-pass filter, high-pass filters of DB4 [8] are given in Figure 2. Reconstruction of low-pass filter and high-pass filters of DB4 are given in Figure 3.

Venkatrama Phani Kumar S[1], KVK Kishore[2] and K Hemantha Kumar[3] are with Vignan's Engineering College, Vadlamudi, Guntur, Andhra Pradesh

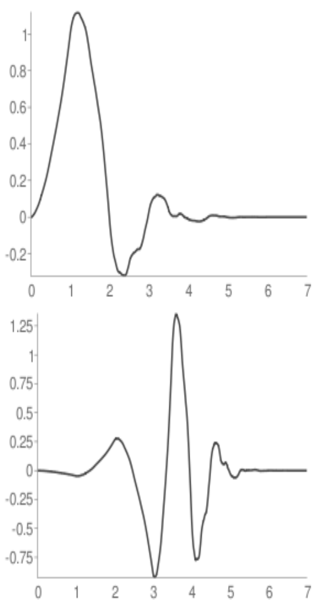


Figure 1: The Scaling & Wavelet functions of DB4

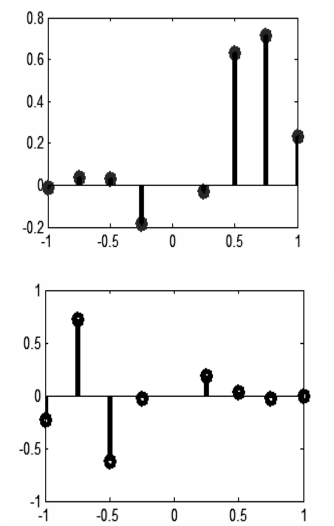


Figure 2: Decompositions of Low-Pass & High-Pass Filters

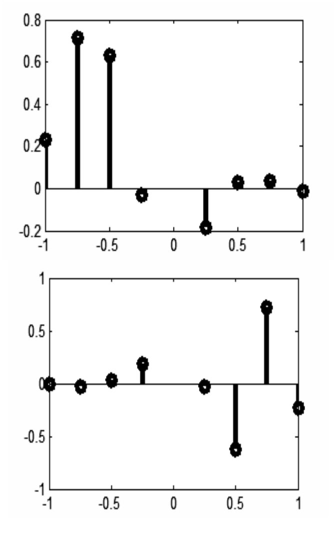


Figure 3: Reconstruction of Low-Pass & High-Pass Filters

II. LOCALLY DISCRIMINATING PROJECTION

In classification problems, the class labels of the training samples are available. The label information can be used to guide the procedure of feature extraction. This is supervised feature extraction, in contrast to the unsupervised feature extraction methods, which take no consideration of the label information. From the theoretical derivation, we can easily find that LPP [5] is an unsupervised feature extraction method. Sine it has little to do with the class information. Some preliminary efforts have already been taken to extend LPP to be a supervised feature extraction method, in which the similarity matrix W can be defined as.

$$W_{ij} = \begin{cases} 1 \text{ if } \mathrm{x_i} \text{and } \mathrm{x_j} \text{belongs to the same class} \\ 0 \end{cases}$$

The basis idea behind it is to make two points become the same point in the feature space, if they belong to the same class. However, this can make the algorithm apt to over fit the training data and sensitive to the noise. Moreover, it can make the neighborhood graph of the input data disconnected. This contradicts the basic idea of LPP. In this situation LPP has close connection to LDA. It means that the manifold structure of the data, which is also very important for classification, which is distorted. To make the LPP algorithm more robust for the classification tasks, LDP [2][1] method is proposed. LDP is the recent one and it is different from PCA and LDA, which aim to preserve the global Euclidean structure, LDP is the extension of LPP. Which preserve the intrinsic geometry structure by learning a locality preserving submanifold? But unlike LPP, LDP is a supervised feature extraction method. LPP is based on the spectral graph theory it is assumed that the n-dimensional subband face (x_1, x_2…..xn) is distributed on a low dimensional submanifold. It is desired to find a set of d dimensional data points (y_1,y_2…..y_N)with the same local neighborhood structure as the (x_1,x_2.. x_n). A weighted graph G= (V, E, W) is constructed, where V is the set of all points, E is the set of edges connecting the points, W is a similarity matrix with weights characterizing the likelihood of two points. Where W can be defined as

$$W_{ij} = \begin{cases} \exp\left(-\frac{\left\|X_i - X_j\right\|^2}{\beta}\right), & \left\|X_i - X_j\right\|^2 < \mathcal{E} \\ 0, & otherwise \end{cases}$$

If x_i is among k nearest neighbors of x_j or x_j is among k nearest neighbors of x_i; the given observation are (x_i, ri) where i=1....N and r_i is the class label of x_i. The discriminating similarity between two points' x_i and x_j is defined as follows.

$$W_{ij} = \begin{cases} \exp\left(-\frac{\left\|X_i - X_j\right\|^2}{\beta}\right)(1 + \exp\left(-\frac{\left\|X_i - X_j\right\|^2}{\beta}\right)) & 1 \\ \exp\left(-\frac{\left\|X_i - X_j\right\|^2}{\beta}\right)(1 - \exp\left(-\frac{\left\|X_i - X_j\right\|^2}{\beta}\right)) & 2 \\ 0 & 3 \end{cases}$$

1- If xi is among k nearest neighbors of x_j or x_j is among k nearest neighbors x_i and $r_i = r_j$.

2- If x_i is among k nearest neighbors of x_j ,x_j is among k nearest neighbors x_i and $r_i \sim= r_j$.

3- otherwise;

Since the Euclidean distance $\|X_i - X_j\|$ is in the exponent, the parameter β is used as a regulator and the selection of β is the problem and it controls the overall scale or smoothing of the space. Very low value for β implies that Wij in above equation will be near zero for all but the closest points. Low value for β therefore minimizes the connections between the points. Very high value for β implies that the Wij is near that Wij in equation (1). Then the equation (2) can be viewed as generalization of the equation (1). Usually the value of the parameter β can be chosen as the square of the average Euclidean distance between all pairs of the data points. In equation (2) the discriminating similarities of the points in the same class are larger than those in LPP whereas the discriminating similarities of the points in different classes are less than those in LPP. This is more favorable for the classification tasks.

Let $\exp\left(-\frac{\|x_i - x_j\|^2}{\beta}\right)$ be the local weight and let $(1+\exp\left(-\frac{\|x_i - x_j\|^2}{\beta}\right))$ and $(1-\exp\left(-\frac{\|x_i - x_j\|^2}{\beta}\right))$ be the intra class discriminating weight and inter class discriminating weight, respectively. Then the discriminating similarity can be viewed as the integration of the local weight and the discriminating weight. It means that the discriminating similarity reflects both the local neighborhood structure and the class information of the data set. As for the discriminating similarity, its properties and the corresponding advantages can be summarized as follows.

Property 1: when the Euclidean distance is equal, the intra class similarity is larger than the interclass similarity. This gives a certain chance to the points in the same class to be 'more similar 'i.e., to have a large value of similarity, than those in different classes. This is suitable for classification tasks.

Property 2: since $1 <= (1+\exp\left(-\frac{\|x_i - x_j\|^2}{\beta}\right)) <= 2$ and $0 <= (1-\exp\left(-\frac{\|x_i - x_j\|^2}{\beta}\right)) < =1$ Thus no matter how strong the noise, the intra class and inter class similarity can be controlled in certain ranges, respectively. This prevents the neighborhood relationship from being forcefully distorted and the main geometric structure of the data set can largely preserved.

Property 3: with the decreasing of the Euclidean distance, the inter class discriminating weight decreases toward 0. It means close points from different classes should have a smaller value of similarity. Thus the margin between different classes becomes larger than that in LPP [5]. On the other hand with the increasing of the Euclidean distance, the local weight decreases toward 0. This endows the discriminating similarity with the ability to prevent the noise. i.e., the more distinct points from the same class should be less similar to each other. Both aspects indicate that the discriminating similarity can strengthen the power of margin augmentation and noise suppression.

Because of these good properties, the discriminating similarity can be used in classification tasks. Since the discriminating similarity integrates both the local information and label information in this novel algorithm called LDP. LDP can map the data into a low dimensional space where points belonging to the same class are close to each other while those belonging to different classes are far away from each other. At the same time, the main manifold structure of the original data can be preserved. Moreover LDP can limit the effect of the noise and augment the margin between different classes. Thus LDP can be used to design a robust classification method for real world problems. To summarize the classification has four steps as follows.

1. Construct the discriminating similarity of any two data points. For each sample x_i set the similarity

$$W_{ij} = W_{ji} = e^{\left(-\frac{\|x_i - x_j\|^2}{\beta}\right)} (1+e^{\left(-\frac{\|x_i - x_j\|^2}{\beta}\right)})$$

if x_j is among the k-nearest neighbors of x_i and $r_i = r_j$; Set the

Similarity $W_{ij} = W_{ji} = e^{\left(-\frac{\|x_i - x_j\|^2}{\beta}\right)} (1-e^{\left(-\frac{\|x_i - x_j\|^2}{\beta}\right)})$

if x_j is among the k-nearest neighbors of x_i and $r_i \sim= r_j$.

2. Obtain the diagonal matrix D and the Laplacian matrix L.

3. Optimize $\varphi^* = \underset{(\varphi^T X D X^T \varphi)=1}{\operatorname{argmin}} \; trace\,(\varphi^T X L X^T \varphi)$.

This can be done by solving the generalized eigenvalue problem $XLX^T\varphi = \lambda XDX^T\varphi$

4. Map the given query using the projection matrix φ* and then predict its class label using a given classifier.

III. Wavelet Based Kernel Locally Discriminating Projection

The proposed face recognition system consists of five phases, (1) Face Detection (2) Face Normalization (3) Image Compression (4) Feature Extraction with W-KLDP (5) Classification. In order to reduce the computational complexity, all the face images are cropped. The main objective of this paper is that existing feature extraction methods like PCA, LDA [7] are used to generate linear features with 70-85% and more discrimination. Achievement of 100% of discrimination to the generated feature vectors of the face images is very difficult. This is due to the similarity that exists among features of the images stored in the database. They possess small variations and may not be able to produce feature vectors with required amount of discrimination. The architecture of the Proposed Method is shown in figure 4.

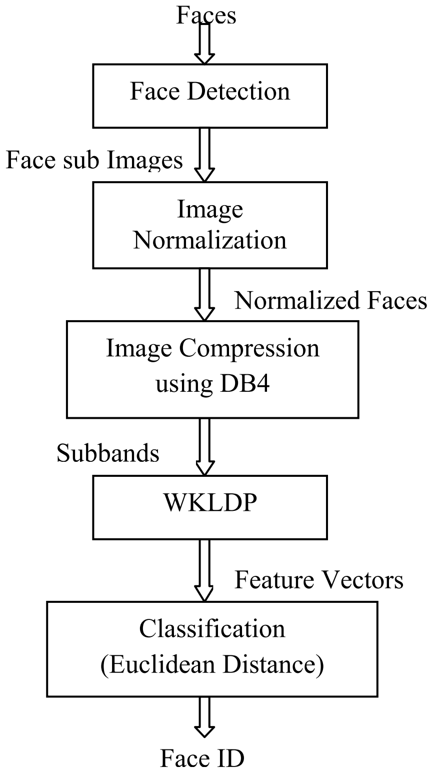


Figure 4: Architecture of Proposed Method

Suppose that the Euclidean space R^n is mapped to a Hilbert space H through a nonlinear mapping function $\varphi: R^n \rightarrow H$. Let $\varphi(x)$ denotes the data matrix in the Hilbert space, $\varphi(X) = [\varphi(x_1), \varphi(x_2), \ldots \varphi(x_m)]$. Now the Eigen vector problem in the Hilbert space can be written as follows: $[\varphi(X)L\varphi^T(X)\}]v = \lambda[\varphi(X)D\varphi^T(X)]v$

To generalize LDP [2] to the nonlinear case, we formulate it in a way that uses dot product exclusively. Therefore, we consider an expression of dot product on the Hilbert space H given by the following kernel function:

$$K(x_i, x_j)=(\varphi(x_i).\varphi(x_j))=\varphi^T(x_i)\varphi(x_j) \quad (1)$$

Because the Eigen vectors of (2) are linear combinations of $\varphi(x_1), \varphi(x_2), \ldots \varphi(x_m)$, there exist coefficients $\alpha_i, i = 1,2,\ldots,m$ such that

$$V = \sum_{i=1}^{m} \alpha_i \varphi(x_i) = \varphi(X)\alpha \quad (2)$$

Where $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_m]^T \in R^m$. By simple algebra formulation, we can finally obtain the following eigenvector problem

$$KLK\alpha = \lambda KDK\alpha \quad (3)$$

Let the column vectors $\alpha^1, \alpha^2, \ldots \alpha^m$ be the solutions of above equation. For a test point x, we compute projections onto the eigenvectors v^k according to

$$((v^k.\varphi(x))=\sum_{i=1}^{m} \alpha_i{}^k(\varphi(x).\varphi(x_i)) = \sum_{i=1}^{m} \alpha_i{}^k K(x, x_i)$$

where $\alpha_i{}^k$ is the i[th] element of the vector α^k. For the original training points the maps can be obtained by y=Kα, where the i[th] element of y is the one dimensional representation of x_i. Furthermore equation (3) can be reduced to

$$Ly = \lambda Dy \quad (4)$$

This is identical to the eigenvalue problem of Laplacian Eigenmaps in equation (2). We normalize the solutions belonging to nonzero eigenvalues by requiring that the corresponding vectors in H be normalized, $(V^k.V^k) = 1$. By virtue of the above equations this translates to

$$1=\sum_{i,j=1}^{l} \alpha_i{}^k \alpha_j{}^k \left(\varphi(x_i).\varphi(x_j)\right) = (\alpha^k.K\alpha^k) = \lambda_k(\alpha^k.\alpha^k)$$

For principal component extraction, we compute projections of the image of a test point φ(x) onto the eigenvectors V^k in H according to

$$(V^k.\varphi(x)) = \sum_{i=1}^{l} \alpha_i{}^k \left(\varphi(x_i).\varphi(x)\right)$$

Therefore we are able to use kernel functions for computing these dot products without actually performing the map φ into some dot product space H. Kernels which have successfully been used in support vector machines include polynomial kernels.

$k(x,y)=(x.y)^d$ radial basis $k(x,y)=\exp(-\left(\frac{\|x-y\|^2}{2\sigma^2}\right))$, and sigmoid kernels functions [9] as follows.$(x,y)=\tanh(k(x.y)+\theta)$. It can be shown that polynomial kernels of degree d correspond to a map φ into a feature space which is spanned by all products of d entries of an input pattern, in this case N=2,d=2.

$$(x.y)^2=(x_1{}^2, x_1x_2, x_2x_1, x_2{}^2)\,(y_1{}^2, y_1y_2, y_2y_1, y_2{}^2)^T$$

If the patterns are images, we can thus work in the space of all products of d pixels and there by take into account higher -order statistics when doing LDP. Substituting kernel functions for all occurrences of (φ(x), φ(y)), we obtain the following algorithm for kernel LDP; we compute the dot product matrix $K_{ij} = (k(x_i, x_j))_{ij}$, solved by diagonalizing K, normalize the components of test point x by computing projections onto eigenvectors.

IV. Experimental Results

The ORL face database consists of ten different images of each of 40 distinct subjects with 92*112 pixels. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open /closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The ten sample face images of a person in shown in figure 5.

Figure 5: Sample ten face images of ORL database

The normalized face images using Gamma Correction are shown in figure 6.

Figure 6: Normalized Face Images

After the first level decomposition with DB4 the size of the ORL face images is decreased to 49*59 and in the second compression the size of the images decreases to 28*33pixels. The sample face images of first and second decompositions are shown in figure 7.

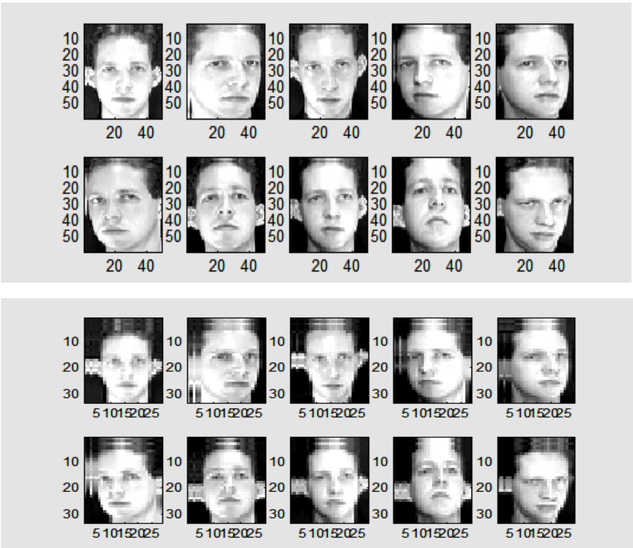


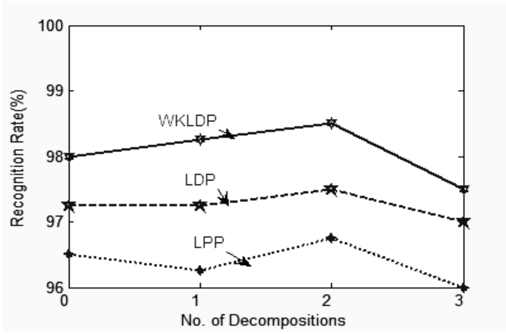
Figure 7: First and Second Decomposed Face Images With DB4

In the experiment, we have tested the recognition rates of LPP, LDP, and WKLDP. These three methods are used for feature extraction, after feature extraction classification is done with Standard Euclidean distance. Totally 6 images of each 40 distinct objects are considered for training and remaining 4 images of 40 distinct objects are used for testing. True Positive is the ratio of correctly identifying the authorized persons.

The recognition rates of LPP, LDP, and WKLDP are given in Table 1. In plot 1 number of decompositions is taken on X axis and recognition rate is taken on Y axis. In all these three the proposed method has proved high recognition capability. Without using of wavelets the time taken for training is so high hence manually all the face images are resized to 40*49 and then KLDP is used.

TABLE 1: TRUE POSITIVE

Algorithm	0	1	2	3
LPP	96.5	96.25	96.75	96
LDP	97.25	97.25	97.5	97
WKLDP	98	98.25	98.5	97.5



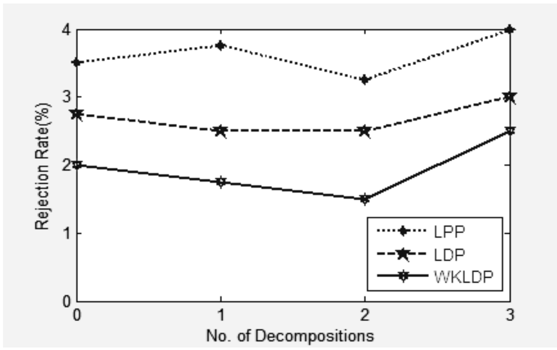
Plot 1: True Positive

The elapsed time taken for training is 1.55 minutes, the training time for first decomposed face images is 2.68 minutes, and for second and third decompositions is 1.05 minutes and 40 seconds respectively. Hence the proposed method has lower time complexity.

Face recognition system failing to verify or identify an authorized person and it also referred as a type I error. FRR is stated as the ratio of the number of false rejections divided by the number of identification attempts. Number of decompositions is taken on X axis and Rejection rate is taken on Y axis. And the results shown in the table 2 gives the proposed method has lower rejection rates of authorized persons.

TABLE 2: FALSE REJECTION RATE

Algorithm	0	1	2	3
LPP	3.5	3.75	3.25	4
LDP	2.75	2.5	2.5	3
WKLDP	2	1.75	1.5	2.5

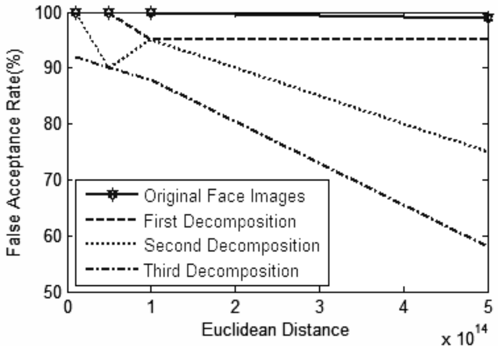


Plot 2: Plot for FRR

Face recognition system incorrectly verify or identify an unauthorized person. It also referred to as a type II error. FAR is stated as the ratio of the number of false acceptances divided by the number of identification attempts. Euclidean Distance is taken on X axis and by varying the Euclidean distance the unauthorized persons (from Indian Face Database) are specified as incorrectly verified means they are shown as unauthorized. And those acceptance rates with different Euclidean distances are given in table 3.

TABLE 3: FALSE ACCEPTANCE RATE

Euclidean Distance	0	1	2	3
1e+013	100	100	100	99
5e+013	100	100	95	95
1e+014	100	90	95	75
5e+014	92	90	88	58



Plot 3: Plot for FAR

V. Conclusion and Further Work

In this paper WKLDP feature extraction method is applied on the subband faces because LDP is failed to extract non linear features of data in a non linear structured space. In the proposed method wavelets are used for dimensionality reduction and to reduce the time complexity in training of faces. The result of subband face images has the features of multi resolution, locality, Sparsity and Decorrelation. These subband faces are used for feature extraction using WKLDP. The experiments are conducted on the ORL face database and the results shown that the proposed method has better performance with lower time complexity.

To improve the recognition capability for unconstrained faces modular LDP combined with wavelets to be proposed in future.

References

[1] Hybrid Face Recognition using Locally Discriminating Projection. Venkatrama Phani Kumar S, K.V.K. Kishore, K Hemantha Kumar. IEEE Conf. ICSAP 2010, pp 327-331

[2] Face Recognition using Wavelet Transform and Locally Discriminating Projection Haitao Zhao, Shaoyuan Sun, Zhongliang Jing, Jingyu Yang Science direct (2006) 1546-1550.

[3] X. He et al, “Face recognition using Laplacian faces”, IEEE Trans. PAMI 27(3) pp 328-340, 2005.

[4] C.Garcia, G.Zikos, and G.Tziritas, “Wavelet packet analysis for face recognition”, Image Vision Computing vol 18, pp 289-297, 2000.

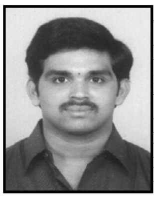
[5] Locality Preserving Projections Xiaofei He, Partha Niyogi.

[6] Statistical and Computational Analysis of Locality Preserving Projection: Xiaofei He, Deng Cai, Wanli Min.

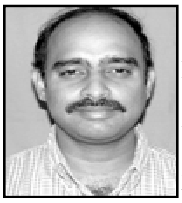
[7] Belhumeur, .P, Hespanha, & Kriegman.D, "Eigenfaces vs.Fisherfaces: Recognition using Class Specific Linear Projection".IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19, No.7, (July1997) 711-720, 0162-8828.

[8] PyWavelets - Discrete Wavelet Transform in Python

[9] J. Yang, A.F. Frangi, J.Y. Yang, et al. Kernel PCA plus LDA: complete kernel Fisher discriminant framework for feature extraction and representation. IEEE Trans. Pattern Anal. Mach Intell, 27(2):230-244, 2005.



Venkatrama Phani kumar S did his Post graduation in Mathematics in Acharya Nagarjuna University in 2007. He is currently working towards M.Tech CSE Project in Vignan’s Engineering College under the supervision of Prof. KVK Kishore and Asso. Prof K Hemantha Kumar. His research interests include Digital image processing, Pattern Recognition and numerical linear algebra.



Prof K.V.Krishna Kishore received his M.Tech degrees in Computer Science and Engineering from Andhra University, Visakhapatnam. He is currently working as Professor and Head of Computer Science and Engineering Department. He got more than 17 years of teaching and Research experience. His current research interests include Digital image processing, Biometric Security, Neural Networks, Pattern Recognition and Medical Imaging.

Mr. K. Hemantha Kumar received his M.Tech degrees in Computer Science and Engineering from Andhra University, Visakhapatnam. He is currently working as Associate Professor in Computer Science and Engineering Department. He got more than 10 years of teaching experience. His current research interests include Digital image processing, Biometric Security, Pattern Recognition.