Journal of Computer Science 5 (11): 801-810, 2009
ISSN 1549-3636
© 2009 Science Publications

# Boosting Kernel Discriminative Common Vectors for Face Recognition

[1]C. Lakshmi and [2]M. Ponnavaikko
[1]Department of Computer Science and Engineering,
SRM University, Kattankulathur, Chennai-603 203, Tamilnadu, India
[2]Bharathidasan University, Trichy, India

**Abstract: Problem statement:** Kernel discriminative common vector (KDCV) was one of the most effective non-linear techniques for feature extraction from high dimensional data including images and text data. **Approach:** This study presented a new algorithm called Boosting Kernel Discriminative Common Vector (BKDCV) to further improve the overall performance of KDCV by integrating the boosting and KDCV techniques. **Results:** In BKDCV, the feature selection and the classifier training were conducted by KDCV and AdaBoost.M2 respectively. To reduce the dependency between classifier outputs and to speed up the learning, each classifier was trained in the different feature space which was obtained by applying KDCV to a small set of hard-to-classify training samples. The proposed method BKDCV possessed several appealing properties. First, like all Kernel methods, it handled non-linearity in a disciplined manner. Second by introducing pair-wise class discriminant information into discriminant criterion, it further increased the classification accuracy. Third, by calculating significant discriminant information, within class scatter space, it also effectively contracted with the small sample size problem. Fourth, it constituted a strong ensemble based KDCV framework by taking advantage of boosting and KDCV techniques. **Conclusion:** This new method was applied on extended yale B face database and achieves better classification accuracy. Experimental results demonstrated the promising performance of the proposed method as compared to the other methods.

**Key words:** Kernel discriminative common vectors, boosting, pair-wise class discriminant information, Adaboost.M2, small sample size problem, discriminant criterion

## INTRODUCTION

Face recognition techniques can be used in a wide range of applications such as identity authentication, access control, military, commercial and surveillance. For these applications, the data are the captured images from a wide variety of sources i.e., sources generating relatively controlled format images and other sources of video images which require additional constraints in terms of speed and processing requirements[1]. The face recognition process includes different stages including localization of faces, feature extraction from the face image, recognition and verification[2].

Research on Human face recognition is going on for a few decades and a number of contributions have already been made[1,3,17]. Among these methods, appearance based approach is one of the most successful and well studied approach for face recognition process[18,28]. This approach operates directly on images or appearance of face objects and processes the images as two dimensional holistic patterns. In these approaches, a two dimensional image of size p by q pixels is represented by a vector in a pq-dimensional space. This space is also called as sample space since face images or samples are represented in this space. Therefore each facial image corresponds to a point in this space and its dimension is very high[4]. This vector point describes the features of a face image. When the number of face images is large, the processing of these images in pq dimensional space is complex and it is very difficult to find the appropriate hyper plane for classification. Thus the dimensionality reduction procedure is required to overcome the problems in high dimensional space. However, since face images have similar structure, the image vectors are correlated and any image in the sample space can be represented in a lower dimensional subspace without losing a significant amount of information.

The eigen face method has been proposed for finding such a lower dimensional subspace[18]. This

**Corresponding Author:** C. Lakshmi, Department of Computer Science and Engineering, SRM University, Kattankulathur, Chennai-603 203, Tamilnadu, India

method uses Principal Component Analysis (PCA), is to find the best set of projection directions in the sample space that will maximize the total scatter across all images. The projection directions are also called the eigenfaces. Any face image in the sample space can be approximated by a linear combination of the significant eigen faces. The sum of the eigen values that correspond to the eigen faces are not used in reconstruction gives the mean square error of reconstruction. This method is an unsupervised technique since it does not consider the classes within the training set data. This approach tends to model unwanted within class variations such as those resulting from the differences in lighting, facial expressions and other factors[5,20]. The criterion used in this method does not attempt to minimize the within class variations, the resulting class tend to have more overlap than other approaches.

The Linear Discriminant Analysis (LDA) also known as Fisher's Linear Discriminant Analysis (FLDA) is proposed in[20]. This method overcomes the limitations of the eigenface method by applying Fisher's linear discriminant criterion[19]. This criterion is used for finding the best set of projection directions in the sample space that will maximize the ratio:

$$J_{FLD}\left(W_{opt}\right) = \arg\max_{W} \frac{\left|W^{T}S_{B}W\right|}{\left|W^{T}S_{W}W\right|}$$

Where:

W = The matrix whose columns are the projection vectors used for feature extraction
$S_W$ = The within class scatter matrix
$S_B$ = The between-class scatter matrix

The above criterion is maximized when the eigen vectors of $S_W^{-1}S_B$ are employed as column vectors in matrix W. Since $S_W^{-1}S_B$ is non symmetric, its eigen decomposition may be unstable. Therefore the major drawback is that it cannot be applied directly since the dimension of the sample space is typically larger than the number of samples in the training set. As a consequence $S_W$ is singular in this case. This problem is known as the "Small Sample Size" (SSS) problem[7].

Various methods have been proposed to solve the SSS problem. Tian *et al.*[29] proposed pseudo inverse method by replacing $S_W$ by its pseudo inverse. In[3,21] the perturbation method is used where a small perturbation matrix is added to $S_W$ in order to make it non singular. However the above methods are computationally expensive since the scatter matrices are very large. Swets and weng[5] proposed a two stage PCA + LDA methods also known as fisher face method, in which PCA is first used for dimension reduction so as to make $S_W$ non singular before the application of LDA. However in order to make $S_W$ nonsingular, some directions corresponding to the small eigen values of $S_T$ are thrown away in the PCA step. Thus there are chances for removing the dimensions that contain discriminative information[6,8,9,30].

A new LDA method was proposed by Chen *et al.*[23] also called null space method which is based on modified fisher's linear discriminant criterion:

$$J_{MFLD}\left(W_{opt}\right) = \arg\max_{W} \frac{\left|W^{T}S_{B}W\right|}{\left|W^{T}S_{T}W\right|}$$

In this method, all image samples are first projected onto the null space of $S_W$, resulting in a new within-class scatter matrix that is a zero matrix. Then PCA is applied to the projected samples to obtain optimal projection vectors. The drawback in this method is they have applied this algorithm only in new reduced space not in the original sample space. But this method performance depends on the null space of $S_W$ in turn which depends on the larger sample space. Thus any kind of preprocessing that reduces the original sample space should be avoided. Another novel method is PCA + Null space method was proposed by Hung *et al.*[8] for dealing with the Small Sample Size (SSS) problem. In this method, at first, PCA is applied to remove the null space of $S_T$ and then the optimal projection vectors are found in the remaining lower dimensional space by using the null space method. Although this method use the original sample space, applying PCA and using all eigen vectors corresponding to the non zero eigen values make these methods impractical for face recognition applications when the training set size is large. This is due to the fact that the computational expense of training becomes very large.

Another method, the Direct-LDA method is proposed in[6]. This method uses the simultaneous diagnolization method[7]. First the null space of $S_B$ is removed and then the projection vectors that minimize the within-class scatter in the transformed space are selected from the range space of $S_B$. However removing the null space of $S_B$ will also remove part of the null space of $S_W$. This may result in the loss of important discriminative information[8,9,22].

Another novel method is Discriminative Common Vector (DCV) method is proposed in[10] which addresses the limitations of above methods for solving SSS problem and also for finding optimal orthonormal projection vectors in the optimal discriminant subspace.

Two efficient algorithms were given to compute the optimal projection vectors. One algorithm uses range space of $S_W$, while the other uses subspace methods and the Gram-Schmidt orthogonalization procedure. However this method can be applied only in the small sample size case and the dimensionality of the null space of the within-class scatter matrix must be large in comparison with the training set size for good recognition rates. Another limitation is that this method extracts only linear features of the samples from the original sample space and it is failed to extract nonlinear features which describe the complexity of face image due to illumination, facial expressions and pose variations[15,17].

The kernel Discriminative Common Vector (KDCV) method is proposed in[24]. This method overcomes the limitations of the DCV method by non-linearly map the original sample space to an implicit higher dimensional feature space. Then the optimal projection vectors are computed in this transformed space. The kernel trick[11] used in this method is an efficient way of nonlinear mapping. Thus the nonlinear features due to illumination, facial expressions and pose variations can be extracted. This method yields an optimal solution for maximizing a modified fisher's linear discriminant criterion. However the performance of KDCV method degrades due to the following non balanced problems. The first problem is, in KDCV method the optimal criterion is based on the conventional between-class scatter matrix which is not directly related to classification accuracy. In particular, the followed dimensionality reduction procedure tends to overemphasize the between class scatter $S_B$ of well separated outlier classes in the sample space at the expense of classes that are close to each other, leading to significant overlap between them. The second problem is the expression of the average within-class scatter has an assumption that all classes have same weight for the covariance. In fact if the class with dominant covariance is an outlier class in the sample space, the within-class scatter $S_W$ will fail to estimate the correct value for improved classification due to this assumption.

In this study we propose a novel KDCV algorithm called Boosting Kernel Discriminative Common Vector algorithm (BKDCV) to overcome the limitations of KDCV. The proposed approach effectively integrates the boosting techniques with the KDCV algorithm based on the pair wise class discriminant information. This BKDCV approach employs the boosting technique to robustly adjust the information and calculate the pair wise class discriminant information that is integrated into the scatter matrices of KDCV in order to solve the non balanced problems in KDCV.

## MATERIALS AND METHODS

**A review of kernel discriminative common vector method:** Assuming that in a set $X=\left\{\left\{x_1^1,x_2^1,...x_N^1\right\},\left\{x_1^2,x_2^2,...x_N^2\right\},...\left\{x_1^c,x_2^c,...x_N^c\right\}\right\}$ there are C classes and each class contains N samples. Let $x_m^i$ be the $m^{th}$ sample in $i^{th}$ class. There are a total of M = N*C samples in the training set X.

In kernel method[12,13,24] the training samples in X are transformed into an implicit higher dimensional feature space F through a non linear mapping function Φ. This mapping function map two vectors that are linearly dependent in the original sample space onto two vectors that are linearly independent in high dimensional feature space F[31]. The feature vector in F is to be computed by computing the inner product of two vectors in F with a kernel function $k(x, y) = \Phi(x)^T \Phi(y)$.

Let the sample matrix set $X=\left\{\left\{x_1^1,x_2^1,...x_N^1\right\},\left\{x_1^2,x_2^2,...x_N^2\right\},...\left\{x_1^c,x_2^c,...x_N^c\right\}\right\}$ which becomes $X_\phi=[\phi(x_1),\phi(x_2),.....\phi(x_M)]$ after samples are mapped into feature space F through a non-linear mapping function. In[15] researchers proved that the relation between the sample matrix $X_\phi$ and the orthonormal basis of the range space of $X_\phi$ in feature space F is based on the fact that the QR decomposition can be derived from the Gram-Schmidt orthogonalization procedure. Then performing the Gram-Schmidt orthogonalization in feature space is equivalent to performing a Cholesky decomposition of the kernel matrix K is also proved in[15]. By using the above concepts in[15] the transformed sample matrix $X_\phi$ can be expressed as:

$$X_\phi = QR \quad (1)$$

Where:
Q = Contains all the basis vectors in feature space F
R = an upper triangular matrix of Q

Formula (1) is the QR decomposition of $X_\phi$ in feature space. Because the columns of matrix Q are orthonormal, the kernel matrix:

$$K=X_\phi^T X_\phi=R^T Q^T QR=R^T R \quad (2)$$

where, K is an n × n kernel matrix which can be computed using kernel function as $(K)_{ij}=k(x_i,x_j)$ and K is a symmetric positive semi definite matrix. From (2) the matrix R can be obtained by performing

Cholesky decomposition of K. Therefore in (1), the relation between the vectors $\phi(x_1)\ldots(x_M)$ and the orthonormal basis vectors $q_1,q_2,\ldots,q_M$ is built and it can be written as:

$$X_\phi R^{-1} = Q \tag{3}$$

The above step shows that performing the Gram-Schmidt orthogonalization in feature space is actually equivalent to performing a Cholesky decomposition of the kernel matrix K[15].

The within class scatter matrix $S_w^\phi$, the between class scatter matrix $S_b^\phi$ and the total scatter matrix $S_t^\phi$ are defined as:

$$S_w^\phi = \frac{1}{N}\sum_{i=1}^{C}\sum_{m=1}^{N_i}\phi(x_m^i)-\mu_i^\phi)\,((\phi(x_m^i)-\mu_i^\phi)^T \tag{4}$$

$$S_b^\phi = \frac{1}{N}\sum_{i=1}^{C} N_i(\mu_i^\phi-\mu^\phi)(\mu_i^\phi-\mu^\phi)^T \tag{5}$$

$$S_t^\phi = S_w^\phi + S_b^\phi \tag{6}$$

Where:
$\mu^\phi$ = The mean of all samples
$\mu_i^\phi$ = The mean of samples of the $i^{th}$ class

The algorithm for KDCV based on subspace methods and cholesky decomposition of kernel matrix K is summarized as follows.

- In transformed feature space F, construct a complete difference subspace is the range space of matrix Bϕ

$$B_\varphi = \left[\phi(b_1^1)\ldots\phi(b_{N-1}^1),\phi(b_1^2)\ldots\phi(b_{N-1}^2)\ldots\phi(b_1^c)\ldots\phi(b_{N-1}^c)\right]$$

Where:

$$\phi(b_k^i) = \phi(b_{k+1}^i)-\phi(x_k^i), k = 1\ldots N$$

Suppose that all orthonormal basis vectors of the sub space of $B_\phi$ form a matrix $Q_B$ and $R_B$ is an upper triangular matrix of $Q_B$.
From (1) we have:

$$B_\phi = Q_B R_B, B_\phi R_B^{-1} = Q_B \tag{7}$$

- The common vectors of each class in feature space F are obtained by projecting any sample from each class onto orthogonal complement of range space of $B_\phi$ as follows:

$$\phi(x_{com}^i) = \phi(x_m^i) - Q_B Q_B^T \phi(x_m^i) \tag{8}$$

The common vector $\phi(x_{com}^i)$ is independent of sample $x_m^i$ of class i. In the following we choose first sample $x_1^i$ from class i

- Find the difference vectors $\phi(b_{com}^k)$ to form matrix $B_{com}^\phi$:

$$B_{com}^\phi = \left[\varphi(b_{com}^1),(b_{com}^2),\ldots(b_{com}^{c-1})\right]$$

Where

$$\phi(b_{com}^k) = \phi(x_{com}^{k+1}) - (x_{com}^1), k = 1\ldots.c-1$$

- Apply Gram-Schmidt orthogonalization procedure to obtain an orthonormal basis $W_{\phi1}$, $W_{\phi1},\ldots,W_{\phi c-1}$ for the range space of $B_{com}^\phi$ which needs to compute kernel matrix $k_{com} = B_{com}^{\phi T}\ B_{com}^\phi$ as follows:

$$(K_{com})_{ij} = \left\langle\phi(b_{com}^i),\phi(b_{com}^j)\right\rangle \tag{9}$$

The upper triangular matrix $R_{com}$ can be obtained by performing the Cholesky decomposition of $k_{com}$ and $k_{com} = R_{com}^T R_{com}$

- According to (3) the optimal projection matrix:

$$W_\phi = \left[(w_{\phi1}, w_{\phi2},\ldots w_{c-1})\right]$$

in feature space F is obtained as

$$W_\phi = \left[(w_{\phi1}, w_{\phi2},\ldots w_{c-1})\right] = B_{com}^\phi R_{com}^{-1} \tag{10}$$

This optimal projection matrix $W_\phi$ maximize the fisher discriminant criteria in feature space F:

$$J(W_{\phi opt}) = \arg\max_{|W_v^T S_w^\phi W_\phi|=0} \left|W_\phi^T S_b^\phi W_\phi\right| \tag{11}$$

where, $S_b^\phi$, $S_w^\phi$ are between class and within class scatter matrix in feature space F

- The discriminative common vector $\Omega_i^\phi$ of class i can be calculated as:

$$\Omega_i^{\phi} = W_{\phi}^{T}\ \phi(x_1^i) = R_{com}^{-1} B_{com}^{\phi} \phi(x_1^i) \quad (12)$$

$\Omega_i^{\phi}$ is independent of sample of class i

Thus KDCV method classifies a new test image $x_{test}$ to class C by finding the minimum Euclidean distance between $\Omega_{test}^{\phi}$ and $\Omega_i^{\phi}$ i.e.:

$$\hat{C} = \min_i \left\| \Omega_i^{\phi} - \Omega_{test}^{\phi} \right\|, i = 1...c \quad (13)$$

**Boosting technique:** Boosting is a general machine learning meta-algorithm for improving the accuracy of any given learning algorithm. One of the most effective boosting algorithms, referred to as AdaBoost, can be used in conjunction with many learning algorithms to improve their performance[14,16]. The AdaBoost algorithm is based on the sample distribution, which measure the hardness of classification of samples. Moreover AdaBoost.M2 algorithm is more suitable for classification of samples in multiclass environment than AdaBoost.M1[16,27]. In this study we prefer using AdaBoost.M2 algorithm in order to effectively overcome the non-balanced problems of KDCV and form a strong connection between KDCV and AdaBoost.M2. The reason for the strong connection is the KDCV method achieves better results on face recognition tasks by means of extracting the non linear features of samples as described in section 2. On the other hand KDCV method suffers from the non balanced problems described in section 1. In our approach we combine the strength of the AdaBoost.M2 algorithm and KDCV method to solve the non balanced problems of KDCV and pair wise class discriminant distribution is introduced on the basis of mislabels distribution from AdaBoost.M2[25] and also it is used to compute the weighted scatter matrices of $S_b^{\phi}$ and $S_w^{\phi}$ which will overcome the problems of KDCV.

The parameters, the pair-wise class discriminant distribution $d_{i,j}$, the relevance based weight for class i, $r_i$, the hardness of separating a sample of class from other classes are used to modify the between class scatter matrix $S_b^{\phi}$ and within class scatter matrix $S_w^{\phi}$ in order to solve the problems of KDCV. Thus these parameters are called boosting parameters and the modified scatter matrices are called weighted scatter matrices. In Adaboost.M2 algorithm, at the $t^{th}$ iteration, the pair-wise class discriminant distribution $d_{i,j}^t$ between classes $C_i$ and Cj can be calculated as:

$$d_{i,j}^t = \begin{cases} \frac{1}{2}\left(\sum_{m=1}^{N_i} \Gamma^t(x_m^i, j) + \sum_{m=1}^{N_j} \Gamma^t(x_m^j, i)\right) & \text{, if } i \neq j \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

Here the mislabel distribution $\Gamma^t$ (*, ●) measures the extent of difficulty of discriminating the example * from the improper label ● on the basis of previous boosting results. A larger value of $d_{i,j}^t$ indicates that the worse separability between class i and class j further embodying also that the class i and j are closer together in F. By using (14), the relevance based weight $r_i^t$ for class i at $t^{th}$ iteration can be calculated which uses the parameter $d_{i,j}^t$ i.e.:

$$r_i^t = \sum_{j \neq i} w(d_{i,j}^t) \quad (15)$$

For simplicity, we let W (*) =* in this study. Larger value of $r_i^t$ shows that the worse separability of class i from other classes at $t^{th}$ iteration.

The other parameter $q_{ii_m}^t$ which represents the difficulty of separating an $m^{th}$ sample of class i from other classes at $t^{th}$ iteration can be calculated by using:

$$q_{ii_m}^t = w\left(\sum_{j \neq i} \Gamma^t(x_{i_m}^i, j)\right) \quad (16)$$

where, $i_m$ = 1,2,…..N and i, j = 1,2,….,C. larger value of $q_{ii_m}^t$.

Shows the worse separability of $m^{th}$ sample of class i from other classes at $t^{th}$ iteration. The reason for the problems of KDCV is the procedure of calculating $S_w^{\phi}$ and $S_b^{\phi}$ does not consider the differences in class variances. This can be modified in present research by using the above computed parameters $q_{ii_m}^t$, $r_i^t$, $d_{i,j}^t$ and generate the weighted scatter matrices $S_B^{\phi}$ and $S_W^{\phi}$. The problems of KDCV can be solved by replacing the scatter operators by means of its weight value. The between class scatter operator $S_b^{\phi}$ is replaced by a weighted between class scatter operator $S_B^{\phi}$:

$$S_B^{\phi} = \sum_{i=1}^{c-1} \sum_{j=i+1}^{c} \frac{N_i N_j}{N^2} w(d_{i,j}^t)(\mu_i^{\phi} - \mu_j^{\phi})(\mu_i^{\phi} - \mu_j^{\phi})^T \quad (17)$$

where, μ is mean of a class. The within class scatter operator $S_W^{\phi}$ is replaced by a weighted within class scatter operator $S_W^{\phi}$:

$$S_W^{\phi} = \frac{1}{N}\sum_{i=1}^{C}\sum_{i_m=1}^{N_i} r_i^t q_{ii_m}^t (\phi(x_{i_m}^i) - \mu_i^{\varphi})(\phi(x_{i_m}^i) - \mu_i^{\phi})^T \quad (18)$$

The problems of KDCV can be addressed by using Eq. 17 and 18. The first problem is addressed based on the pair wise class discriminant distribution and the classes that are not well separated in F is heavily weighted in F. The second problem can be addressed by the parameters relevance based weight $r_i^t$ and the separability of a sample of class from other classes $q_{ii_m}^t$. By using the parameter $r_i^t$ the estimated $S_W^{\phi}$ is only influenced slightly if class $X_i$ is an outlier class and by evaluating the parameter $q_{ii_m}^t$ the analysis of classification of an example $x_{i_m}^i$ can be done with respect to the previous boosting results. In this way the problems of KDCV can be solved by using the above boosting parameters.

**Proposed boosting kernel discriminative common vector algorithm:** Based on Adaboost.M2 algorithm and the weighted scatter operators $S_W^{\phi}$, $S_B^{\phi}$ we propose the modified KDCV algorithm called Boosting Kernel Discriminative Common Vector (BKDCV) algorithm. In BKDCV algorithm, the KDCV technique with weighted scatter matrices is used to extract discriminative common vector of a class in feature space F. In addition, discriminative common vector method is a strong feature extraction technique for classification[10]. As a result, the boosting process cannot go forward due to the very small pseudo loss ε. In general, some sampling procedures are employed to artificially weaken the discriminant technique and in BKDCV, we choose some examples in each class based on $q_{ii_m}^t$ to focus hardest examples in each class. For the features extraction, a simple nearest neighbor classification is generally employed for classification. In this study, we have applied KDCV technique for feature extraction and it generates discriminative common vector of each class in all the iterations. In order to be consistent with the Adaboost Algorithm in our method, the hypothesis $h_t$ (*, ●) between sample and class can be built easily based on KDCV method of classification. The hypothesis constructed $h^t$ is for different hardest sample set in the training set. Thus the discriminative common vector of each class can be generated in a refined manner in its iterations.

To recognize a given test image in BKDCV, the hypothesis $h_t$ built for all the classes in iterations t = 1 to $T_{max}$. The $T_{max}$ represents the maximum number of hardest samples in X. In this study we have considered the worst case of $T_{max}$. Then the given test image is recognized by:

$$h_f(x_{test}) = \arg\max_{i\in\{1...c\}}\left\{\sum_{t=1}^{T_{max}} \log\left(\frac{1}{\beta t}\right) h_t(y^t, i)\right\} \quad (19)$$

where, $h_f(x)$ is the final hypothesis for a given test image. Equation 19 shows the method of assigning a test image to a class which scored maximum response in all the iterations.

**Algorithm:**

Input:-

(i) A set of training examples X

$$X = \left\{ x_{i_m}^i \middle| x_{i_m}^i \in R^n, i = 1...c, i_m = 1...N_i \right\}$$

(ii) A set of all mislabels M

$M = \{[(i, i_m), j] \mid i, j \in \{1, 2,\ldots C\}, i_m \in \{1,\ldots N_i\}, i \neq j\}$

(iii) The initial Mislabel Distribution on M is:

$$\Gamma^t(x_{i_m}^i, j) = \frac{1}{N(C-1)} = \varepsilon$$

a small constant

**Procedure:**

Let $T_{max} = C(N-1)$:

(i) For t = 1.... $T_{max}$ do

(i) Calculate the terms $d_{i,j}^t$ by:

$$d_{ij}^t = \begin{cases} \frac{1}{2}\left(\sum_{i_m=1}^{N_i} \Gamma^t(x_{i_m}^i, j) + \sum_{j_m=1}^{N_j} \Gamma^t(x_{j_m}^j, i)\right), & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

This step shows the separability between class i and j.

(ii) Calculate the relevance based weight for class i, $r_i^t$:

$$r_i^t = \sum_{j\neq i} w(d^t{}_{i,j})$$

This step shows the separation of class i from other classes.

(iii) Calculate the separability of sample $i_m$ of class i from other classes $q_{ii_m}^t$ :

$$q_{ii_m}^t = w\left(\sum_{j\neq i}\Gamma^t(x_{i_m}^i, j)\right)$$

(iv) Select S hardest examples per class based on $q_{ii_m}^t$ to form a training subset $T_s \subset X$. Larger value of $q_{ii_m}^t$ shows hardest samples of class i.

(v) Compute the modified boosting between class scatter matrix $S_B^\phi$:

$$S_B^\phi = \sum_{i=1}^{c-1}\sum_{j=i+1}^{c}\frac{N_i N_j}{N^2} w(d_{i,j}^t)(\mu_i^\phi - \mu_j^\phi)(\mu_i^\phi - \mu_j^\phi)^T$$

And the modified boosting within class scatter matrix $S_W^\phi$:

$$S_W^\phi = \frac{1}{N}\sum_{i=1}^{c}\sum_{i_m=1}^{N_i} r_i^t q_{ii_m}^t(\phi(x_{i_m}^i) - \mu_i^\phi)(\phi(x_{i_m}^i) - \mu_i^\phi)^T$$

(vi) apply KDCV in section2 on $T_s$ and constitute the KDCV based feature extraction technique, denoted by $KDCV_t$.

(vi) Apply $KDCV_t$ on $X_\phi$ and form $Y^t$:

$$Y_t = \left\{y_{i_m}^{i,t} \in R^r\right\}$$

where, $Y^t$ consists of common vector of all the classes at $i^{th}$ iteration.

(vii) Built the hypotheses $h_t(\text{Sample, class}) \in [0,1]$ on the subset $Y^t$ corresponding to $T_s$

(viii) Calculate the pseudo loss based on $h_t$ as:

$$\varepsilon_t = \frac{\sum_{[i,i_m)j]\in M}\Gamma^t(x_{i_m}^i, j)(1 + h_t(y_{i_m}^{i,t}, j) - h_t(y_{i_m}^{i,t}, i))}{2}$$

(ix) Set $\beta_t = \varepsilon_t / (1-\varepsilon_t)$ and

If $\beta_t \leq \varepsilon$ ie., $\frac{1}{N(C-1)}$ then $T_{max} = t - 1$ and break.

(x) Update the mislabel distribution $\Gamma^t$:

$$\Gamma^{t+1}(x_{i_m}^i, j) = \Gamma^t(x_{i_m}^i, j))\beta_t^{(1+h_t(y_{i_m}^{i,t}i) - h_t(y_{i_m}^{i,t}j))/2}$$

(xi) Normalize $\Gamma^{t+1}$:

$$\Gamma^{t+1}(x_{i_m}^i, j) = \Gamma^{t+1}(x_{i_m}^i, j) / (\sum_{[(l,l_m),g]\in M}\Gamma^{t+1}(x_{l_m}^l, g))$$

end for

**Output:** The final hypothesis $h_f(x)$ is:

$$h_f(x) = \arg\max_{i\in\{1...c\}}\left\{\sum_{t=1}^{T_{max}}\log\left(\frac{1}{\beta t}\right)h_t(y^t, i)\right\}$$

For a given sample x i.e. test sample, $Y^t \in R^r$ in all the iterations is the corresponding non linear feature vector extracted by kernel Discriminative common vector method and the maximum response of class is the class label for a given test sample x.

## RESULTS

The Yale B face databases[32] were used to test our proposed method In this, portion of the Yale B face database were retrieved for our testing. Figure 1 shows the three sample sets of Yale B database. The retrieved Yale B face database consists of images from C = 100 different people, using 10 images from each person, for a total of 1000 images. The image contains variations with the following facial expression as center-light, left-light, normal, right-light, happy, sad, sleepy and surprised. First these images were converted to grayscale images. Second we preprocessed these images by aligning and scaling them so that the distance between the eyes was the same for all images and also ensuring that the eyes occurred in the same coordinates of the image. The resulting images were then cropped. The final size of the images was 92×112. The training set consisted of seven images that were randomly selected from each subject and the rest of the images were used for constructing test set. Thus a training set of 700 images and a test set of 300 images were created. This process was repeated 5 times and 5 different training and test sets were created. These five training set and testing set were created by randomly selecting samples from classes at each trial. These sets are used by all the three methods for training and for testing. On each trial DCV, KDCV and our proposed method BKDCV is applied and found its recognition rate.

Fig. 1: Three sample sets from the Yale B face database

Table 1: Comparison of average recognition rates (%)

| No. of training samples in each class | DCV | KDCV | BKDCV |
|---|---|---|---|
| N = 3 | 90.1 | 91.5 | 92.5 |
| N = 5 | 93.7 | 95.3 | 96.1 |
| N = 7 | 96.4 | 97.6 | 99.3 |

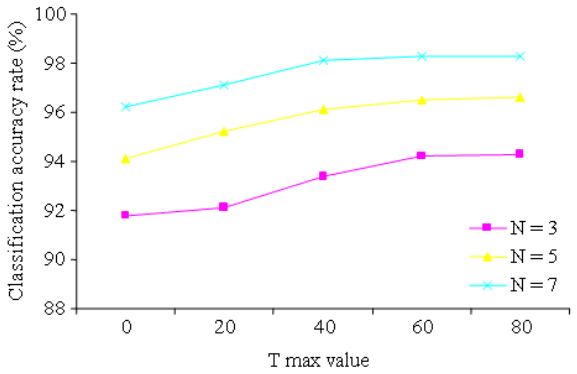


Fig. 2: Comparative performance of BKDCV under different $T_{max}$ values on varying number of samples in a class

The final recognition rate of each method is the average recognition rate of 5 trials in each method. As a result the observed recognition rate of each method on different number of samples in a class is shown in Table 1 and also it shows our proposed method response is better than the other two methods. We observed that in the case of overlapping state the performance of DCV, KDCV method is poor as compared to other methods since the assumption in DCV, KDCV is all classes having the same covariance structure. The Table 1 shows that as the number of samples in the training set increases, the recognition rate also increases. The reason for the improvement in classification accuracy is, the hypothesis constructed on each iteration is considered in the final hypothesis of classification of test image. Another reason is the features which are not considered in iteration are considered in further iterations. Thus the constructed discriminative common vector of a class includes all the features of samples which increase the accuracy.

Another experiment on comparative of the classification accuracy of BKDCV under different $T_{max}$ values on training data sets for varying number of samples in a class is shown in Fig. 2. The experimental results in Fig. 2 shows that the increases in number of samples in the training set increase the classification accuracy. Another observation is that the increasing number of iterations also increases the classification accuracy. At certain state increasing the number of iterations does not react on the classification accuracy due to the method of formation of subset $T_s$ in training time. Because of the classification procedure and the selection of hardest samples for training based on mislabel distribution makes our proposed method works well as compared to other face recognition methods in terms of accuracy.

## DISCUSSION

Accuracy, execution speed are some of factors that may be used for validating the face recognition method. Experimental results show that the proposed method yielded the highest performance in terms of accuracy. Since our method training is based on the hardest samples in the training set, the execution speed of our method is very high as compared to DCV and KDCV methods. The hypothesis constructed $h^t$ is for different hardest sample set in the training set. Thus the discriminative common vector of each class can be generated in a refined manner in its iterations. This work can be extended for face images varying in aging and pose.

## CONCLUSION

In this study, a novel KDCV algorithm has been presented by incorporating the boosting technique into KDCV and called as Boosting Kernel Discriminative Algorithm. The Adaboost.M2 algorithm is used as boosting technique. On each iteration, this algorithm updates the mislabel distribution parameter $\Gamma^t$ based on the previous boosting results. This increases the classification accuracy. This new algorithm BKDCV effectively integrates the strengths of the boosting and KDCV techniques to give an ensemble based KDCV framework with strong nonlinear feature extraction capability and overcomes the problems in KDCV method. The experimental results on Yale B face data base show that the proposed BKDCV method enhances the performance of KDCV method of face recognition process.

## REFERENCES

1. Chellappa, R., C.L. Wilson and S. Sirohey, 1995. Human and machine recognition of faces: A survey. Proc. IEEE, 83: 705-741. DOI: 10.1109/5.381842
2. Zhao, W., R. Chellappa and A. Krishnaswamy, 1998. Discriminant analysis of principal components for face recognition. Proceeding of the 3rd IEEE International Conference Automatic Face and Gesture Recognition, Apr. 14-16, IEEE Computer Society, Washington DC., USA., pp: 336. http://portal.acm.org/citation.cfm?id=796069

3. Zhao, W., R. Chellappa, A. Rosenfeld and P.J. Phillips, 2000. Face recognition: A literature survey. ACM Comput. Surveys, 35: 399-358. http://portal.acm.org/citation.cfm?doid=954339.954342
4. Turk, M., 2001. A random walk through eigenspace. IEICE Trans. Inform. Syst., E84-D: 1586-1695. http://search.ieice.org/bin/summary.php?id=e84-d_12_1586&category=D&year=2001&lang=E&abst=&auth=1
5. Swets, D.L. and J. Weng, 1996. Using discriminant eigenfeatures for image retrieval. IEEE Trans. Patt. Anal. Mach. Intel., 18: 831-836. DOI: 10.1109/34.531802
6. Yu, H. and J. Yang, 2001. A direct LDA algorithm for high- dimensional data with application to face recognition. Patt. Recog., 34: 2067-2070. DOI: 10.1016/S0031-3203(00)00162-X
7. FuKunaga, K., 1990. Introduction to Statistical Pattern Recognition. 2nd Edn., Academic Press, New York, USA., ISBN: 10: 0122698517, pp: 295.
8. Huang, R., Q. Liu, H. Lu and S. Ma, 2002. Solving the small size problem of LDA. Proc of the 16th International Conference on Pattern Recognition, Aug. 2002, IEEE Xplore Press, USA., pp: 29-32. DOI: 10.1109/ICPR.2002.1047787
9. Yang, J., D. Zhang and J.Y. Yang, 2003. A generalized K-L expansion method which can deal with small sample size and high- dimensional problems. Patt. Anal. Appli., 6: 47-54. DOI: 10.1007/s10044-002-0177-3
10. Cevikalp, H. M. Neamtu, M. Wilkes and A. Barkana, 2005. Discriminative common vector for face recognition. IEEE Trans. Patt. Anal. Mach. Intel., 27: 4-13. DOI: 10.1109/TPAMI.2005.9
11. Baudat, G. and F. Anouar, 2000. Generalized discriminant analysis using a kernel approach. Neural Comput., 12: 2385-2404. http://www.ncbi.nlm.nih.gov/pubmed/11032039
12. Müller, K.R., S. Mika, G. Rätsch, K. Tsuda and B. Schölkopf, 2001. An introduction to kernel-based learning algorithms. IEEE Trans. Neural Networks, 12: 181-201. DOI: 10.1109/72.914517
13. Shawe-Taylor, J. and N. Cristianini, 2004. Kernel Methods for Pattern Analysis. Cambridge University Press, England, ISBN: 0521813972, pp: 462.
14. Freund, Y. and R.E. Schapire, 1996. Experiments with a new boosting algorithm. Proceeding of the 10th International Conference on Machine Learning, (ML'96), Morgan Kaufmann Publishers, Inc,. USA., pp: 148-156. http://direct.bl.uk/bld/PlaceOrder.do?UIN=018249972&ETOC=RN&from=searchengine
15. He, Y.H., L. Zhao and C.R. Zou, 2005. Kernel discriminative common vectors for face recognition. Proceeding of the 4th International Conference on Machine Learning and Cybernetics, Aug. 18-21, IEEE Xplore Press, Guangzhou, China, pp: 4605-4610. DOI: 10.1109/ICMLC.2005.1527750
16. Dai, G. and D.Y. Yeung, 2007. Boosting kernel discriminant analysis and its application to tissue classification of gene expression data. Proceeding of the 20th International Joint Conference on Artificial Intelligence, (AI'07), pp: 744-749. http://www.aaai.org/Papers/IJCAI/2007/IJCAI07-119.pdf
17. Abate, A.F., M. Nappi, D. Riccio and G. Sabatino, 2007. 2D and 3D face recognition: A survey. Patt. Recog. Lett., 28: 1885-1906. http://portal.acm.org/citation.cfm?id=1283081
18. Turk, M. and A.P. Pentland, 1991. Eigenfaces for recognition. J. Cogn. Neurosci., 3: 71-86. http://portal.acm.org/citation.cfm?id=1326894
19. Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen., 7: 179-188. http://www.citeulike.org/user/walkking/article/764226
20. Belhumeur, P.N., J.P. Heepanha and D.J. Krtegman, 1997. Eigenfaces Vs fisherfaces: Recognition using class specific Kinear projection. IEEE Trans. Patt. Anal. Mach. Intel., 19: 711-720. DOI: 10.1109/34.598228
21. Hong, Z.Q. and J.Y. Yang, 1991. Optimal discriminant plane for asmaller number of samples and design method of classifier on the palne. Patt. Recog., 24: 317-324. DOI: 10.1016/0031-3203(91)90074-F
22. Bing, Y., J. Lianfu and C.O. Ping, 2002. A new LDA based method for face recognition. Proceeding of the 16th International Conference in Pattern Recognition, Aug. 11-15, IEEE Computer Society, Washington DC., USA., pp: 168-171. http://portal.acm.org/citation.cfm?id=842665
23. Chen, L.F., H.Y.M. Liou, MT. Ko, J.C. Lin and G.J. Yu, 2000. A new LDA based face recognition system which can solve SSS problem. Patt. Recog., 33: 1713-1726. DOI: 10.1016/S0031-3203(99)00139-9
24. Cevikalp, H., M. Neamtu and M. Wilkes, 2006. Discriminative common vector method with kernels. IEEE Trans. Neural Networks, 17: 1550-1565. DOI: 10.1109/TNN.2006.881485
25. Freund, Y. and R.E. Schapire, 1995. A decision theoretic generalization of online learning and an application to boosting. Lecturer Notes Comput. Sci., 409: 23-37. DOI: 10.1007/3-540-59119-2_166

26. Lizhao, Y.H., L. Zhao and C.R. Zou,2005. Kernel discriminative common vector for face recognition. Proceeding of the 14th International Conference on Machine Learning and Cybernetics, Aug. 18-21, IEEE Xplore Press, Guangzhou, China, pp: 4605-4610. DOI: 10.1109/ICMLC.2005.1527750
27. Lu, J.W., K.N. Plataniotis and A.N. Venetsanopoulos, 2003. Boosting linear discriminant analysis for face recognition. Proceedings of the IEEE International Conference on Image Processing, Sept. 14-17, IEEE Xplore Press, USA., pp: 657-660. DOI: 10.1109/ICIP.2003.1247047
28. Murase, H. and S.K. Nayar, 1995. Visual learning and recognition of 3-D objects from appearance. Int. J. Comput. Vis., 14: 5-24. http://portal.acm.org/citation.cfm?id=208946.208947
29. Tian, Q., M. Barbero, Z.H. Gu and S.H. Lee, 1986. Image classification by the foley-sammon transform. Opt. Eng., 25: 834-840.
30. Dai, D.Q. and P.C. Yuen, 2003. Regularized discriminant analysis and its application to face recognition. Patt. Recog., 36: 845-847. DOI: 10.1016/S0031-3203(02)00092-4
31. Burges, C.J.C., 1998. A tutorial on support vector machines pattern recognition. Data Min. Know. Dsicovery, 2: 121-67. http://portal.acm.org/citation.cfm?id=593463
32. Georghiades, A.S. and P.N. Behumeur and D.J. Kriegman, 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Patt. Anal. Mach. Intel., 23: 643-660. DOI: 10.1109/34.927464