INTERNATIONAL JOURNAL OF COMPUTATIONAL MATHEMATICAL IDEAS
ISSN:0974-8652 / www.ijcmi.webs.com / VOL 1-NO3-PP 73-79 (2009)

IMPROVED KERNEL COMMON VECTOR METHOD FOR FACE RECOGNITION

C. Lakshmi[#], Dr. M. Ponnavaikko[*]

[#] *Research Scholar, School of Computer Science & Engg., SRM University,* ***e-mail id :slva90@yahoo.co.in***
[*] *Vice Chancellor, Bharathidasan University, Senior IEEE member,* ***e-mail id: ponnav@gmail.com***

Abstract:

The common vector (CV) method is a linear subspace classifier for datasets, such as those arising in image and word recognition. In this approach, a class subspace is modeled from the common features of all samples in the corresponding class. Since the class subspace are modeled as a separate subspace for each class in feature domain, there is overlapping between these subspaces and there is loss of information in the common vector of a class. This reduces the recognition performance. In CV method the followed criterion considers only the class scatter matrices. Thus the neglecting of the influence of neighboring classes in CV also reduces the recognition performance. In multi-class problems, within-class and between-class scatter should be considered in classification criterion. Since the scatter matrices Sw and Sb followed in Discriminative common vector (DCV) based on the assumption that all classes have similar covariance structures, these scatter matrices cannot be followed in CV method. Generally a linear subspace classifier fails to extract the non-linear features of samples which describe the complexity of face image due to illumination, facial expressions and pose variations. In this paper, we propose a new method called "Improved kernel common vector method" which solves the above problems by means of its appealing properties. First the boosting parameters in the proposed between-class and within-class scatter matrices consider the neighboring classes and a sample of a class with other classes increases the recognition performance. This makes the obtained common vector has more significant discriminant information. Second like all kernel methods, it handles non-linearity in a disciplined manner which extracts the non-linear features of samples representing the complexity of face images. Experimental results on real time face database demonstrate the promising performance of the proposed methodology.

Keywords: subspace classifier, kernel common vector, pair wise class discriminant criterion, mislabels distribution, boosting parameter, scatter operators.

1. Introduction:

In the last two decades, face recognition technology has become one of the most wanted technologies for its potential applications in personal identification, security access control, surveillance, telecommunications, digital libraries, human-computer interaction, military and so on [1],[2],[28]. Face recognition can be described as the identification process of a given input face image using a stored face database [3].

The extraction of best representative features from the face images significantly improves the performance of any face recognition system. Thus the feature extraction is a critical issue in face recognition tasks. Numerous approaches have been proposed in recent years. Among them linear subspace analysis is one of the most popular feature extraction methods, which aims to project face images into a low-dimensional subspace with a linear transformation. This linear subspace classifier uses a linear subspace for each class [4]. In this method, it is assumed that the vector distribution of each class corresponds to a lower dimensional subspace of the original sample space. The subspaces representing classes are defined in terms of basis vectors that are linear combinations of the sample vector of each class. Therefore, basis vectors spanning those subspaces must first be computed. Also, determining the dimension of each subspace is a major issue since subspace dimensions have a strong influence on the performance of the subspace classifier. In particular, large subspace dimensions lead a low recognition performance due to the overlapping regions among classes whereas small subspace dimensions increase the error rates because of a poor resulting approximation [5], [6].

The first subspace method [7] is the class-featuring information processing (CCAFIC) for pattern classification. This method employs the principal component analysis (PCA) to compute the basis vectors spanning the subspace of each class. However, the subspaces may sometimes have a large overlapping region in common, which causes poor classification of data samples. Therefore, the method of orthogonal subspace was proposed which attempts to remove the common regions of the classes and makes the class subspaces mutually orthogonal [8]. Another new method proposed in [9] selects the basis vectors in such a way that the projections onto the so-called rival subspaces are minimized.

In some classification tasks, the dimensionality of the sample space can be larger than the number of training samples in each class. The computations that carried out at full dimensionality may not deliver the advantages of high dimensional sample space when the number of training samples is insufficient. Therefore the dimensionality of the sample space is usually reduced before applying a classifier to data samples in the original sample space. The Eigen face method [10] and Fisher's linear discriminant analysis [11] are two popular feature extraction methods that are used for dimension reduction. However, dimensionality reduction through feature extraction may cause loss of important discriminatory information. Therefore [12] proposed the common vector subspace classifier method that models subspaces in such a way that their basis vectors span the null space of the co-variance matrix of each class, assuming that the number of samples in each class is smaller than or equal to the dimensionality of the sample space [13], [12]. This method has been successfully applied in the isolated word recognition and face recognition problems [12], [14]. The class decision boundaries that are obtained by linear subspaces are quadratic. However, the linear subspaces may not extract non-linear features of classes since each class is associated with a linear subspace [15], [16]. The kernel-based non-linear subspace method, which is called the kernel class featuring information compression was developed to overcome the limitation [15],[16]. In this, it is assumed that all samples in each class lie in some non-linear subspace. Therefore, all data samples in each class are mapped to a high dimensional feature space through a non-linear kernel mapping function and the kernel PCA [17] has been employed to compute the principal components of the correlation matrices of classes in the mapped space. This process spreads the data over a greater volume which in turn reduces overlapping regions among the classes. In this, the mapped function or the mapped samples are not used explicitly, which makes this method feasible.

Another new non-linear method, the kernel common vector method was developed in [18]. In this method, the training set samples are mapped into an implicit higher dimensional space using a non-linear kernel mapping function and then applying the linear CV method in the mapped space .In this the kernel functions are used to compute the dot products of the mapped samples. As a result the mapping function and the mapped samples are not used explicitly which makes the method feasible. However, in this method, the followed discriminative criterion considers the within-class scatter only. Within-class scatter matrix may cause overlapping regions between the class subspaces. This makes the criterion neglecting the overlapping between the classes and thus reduces the recognition performance. This shows the importance of consideration of between-class scatters also. Rifat Edizkan et al[19] proposed an improvement in common vector method by introducing two different optimization criterion [18] in their study report. In this, the first optimization criterion, the distances of inter-class scatter are maximized in the indifference subspace obtained from the within-class scatters and in the second optimization criterion, the distances of intra-class scatter are minimized in the difference subspace obtained from the between-class scatters. The obtained class subspace dimension is less as compared to the subspace obtained in KCV method. The results in [19] show among these two criterions the first criterion provides advantages in designing recognizer in processing time and in memory space utilization. However, in first criterion, the methods followed in the distances of inter-class scatter are measured in the indifference subspace leads to overlapping between the class subspaces. The within-class scatter criterions of [18] consider only the class co-variance and neglecting the other class co-variances. Thus the within-class scatter fails to estimate the correct value for improved classification.

In this paper, we propose a new method called "Improved Kernel Common Vector" (IKCV) which integrates the boosting parameters with scatter operators and then applying KCV which works based on our proposed scatter operators. The IKCV employs the boosting parameters to robustly calculate the pairwise class discriminate information that is integrated into the proposed within-class and between-class scatter operators in order to solve the problems in KCV. The boosting techniques has been applied to LDA, KDA, and showed the recognition performance [20], Our proposed method has many advantages as follows: (i) it handles non-linearity in a disciplined manner (ii) by introducing the pair wise class discriminant criterion in the scatter operators improves the discriminant information consider for recognition. This increases the recognition speed with accuracy. The remainder of this paper is organized as follows: In section II, a brief review of the KCV method is presented; section III describes the proposed within-class and between-class scatter operators. In section IV we describe our proposed method "Improved Kernel Common Vector" method. We discuss our experimental results in section V and our conclusions are given in section VI.

2. A brief review of KCV method:

Let the training set be composed of C classes, where the i^th class contains N_i samples and let x_m^i be a d-dimensional column vector, which denotes the m^th sample from the i^th class. There is a total of M= $\sum_{i=1}^{C} N_i$ samples in the training set.

In this method, the training set samples are mapped into an implicit higher dimensional space using a non-linear kernel mapping function and then applying the linear CV method in the mapped space. In high dimensional space, let $\Phi=[\phi^{(1)},\phi^{(2)},....,\phi^{(i)}]$ represents the matrix whose columns are the mapped training samples where $\Phi^{(i)}=[\phi(x_1^i),\phi(x_2^i),....,\phi(x_{N_i}^i)]$ is the matrix whose columns are mapped samples of the i^{th} class.

The scatter matrix S_i^{Φ} of each class is defined as

$$S_i^{\Phi}=\sum_{m=1}^{N_i}\left(\Phi\left(x_m^i\right)-\mu_i^{\phi}\right)\left(\Phi\left(x_m^i\right)-\mu_i^{\phi}\right)^T \quad (1)$$

i=1… C

where μ_i^{Φ} is the mean of samples of the i^{th} class. The total scatter matrix is defined as,

$$S_T^{\phi}=\sum_{i=1}^{C}\sum_{m=1}^{N_i}\left(\Phi\left(x_m^i\right)-\mu^{\phi}\right)\left(\Phi\left(x_m^i\right)-\mu^{\phi}\right) \quad (2)$$

where μ^{Φ} is mean of all samples. The kernel matrix of the mapped data is given as $K=\Phi^T\Phi=(K^{ij})$, i= 1, . ., C, j=1, . . . ,C, where each sub matrix K^{ij} is defined as

$$(K_{mn}^{ij})=<\phi\left(x_m^i\right),\phi\left(x_n^j\right)> \quad (3)$$

$m=1,...,N_i,\ n=1,...,N_j$

The class subspace followed is the intersection of the null space of that class scatter matrix and the range space of the total scatter matrix that is $N\left(S_i^{\phi}\right)\cap R\left(S_T^{\phi}\right)$, i=1,…,C. The basis vectors spanning these spaces represent class subspaces.

The basis vectors spanning these mentioned intersection subspace can be found by using the projection matrices of the null space of a class and range space of total scatter matrix. Since the projection matrices of above spaces are commuted as shown in theorem 1 in [18], the projection matrix $P_{\text{int}}^{(i)}$ of the intersection subspace

$$P_{\text{int}}^{(i)}=P_{NS}^{(i)}P=PP_{NS}^{(i)},\ i=1,\ldots,C \quad (4)$$

Where $P_{NS}^{(i)}$ is the projection matrix of the nullspaces of class i and P is the projection matrix of the subspace of the total scatter matrix.

The basis vectors spanning each intersection subspace $P_{\text{int}}^{(i)}$, i= 1,…,C can be found by transferring the training set samples onto $R\left(S_T^{\phi}\right)$ using the kernel PCA and then compute the null space of the scatter matrix of each class in the transformed space by using Eigen decomposition method.

The result of the above method gives the basis vectors for all the classes projections intersection subspaces ie., the column vector $W^{(i)}$ represent the basis vectors of the intersection subspaces of class i. The transformation of samples onto their corresponding intersection subspace gives the feature vector of the class. This vector is called as CV of class. The transformation of all other samples in this intersection subspace gives the same vector shows the CV is independent of m. The CV of class i is defined as

$$\Omega_{com}^{i}=W^{(i)^T}\phi\left(x_m^i\right),\ i=1,\ldots,C,\ m=1,\ldots,N_i. \quad (5)$$

In this method, the given test image sample x_{test} is recognized by comparing the feature vector of a test sample with the CV of all the classes using Euclidean distance method. Thus the test sample is assigned to the class that minimizes the distance.

3. Proposed within-class and between-class scatter operators:

The common vector is a unique vector which represents the common properties of each class. These common properties represent features that are common to all samples in each class. This vector is used in the recognition of classes in CV method. In multi-class problems, the samples are grouped in the form of classes and the sample space consists of classes. Thus the within-class and between-class scatters should be considered in classification criterion. In this paper, we propose an enhanced within-class and between-class scatter operators in order to address the limitations mentioned in section 1. The within-class scatter of the class will be close to the average of that class. The within-class scatter represents the distance between a sample and other samples in the class. Based on the above concept the within-class scatter matrix is defined as $S_W^{i_\phi}$

$$S_W^{i_\phi}=\sum_{m=1}^{N_i}\left(\Phi\left(x_m^i\right)-\mu_i^{\phi}\right)\left(\Phi\left(x_m^i\right)-\mu_i^{\phi}\right)^T \quad (6)$$

In high dimensional feature space, the distances between the classes are represented by between-class scatter matrix $S_B^{i_\phi}$. The between-class scatter matrix $S_B^{i_\phi}$ is defined as

$$S_B^{i_\phi}=\sum_{i=1}^{C-1}\sum_{j=I+1}^{C}\left(\mu_i^{\phi}-\mu_j^{\phi}\right)\left(\mu_i^{\phi}-\mu_j^{\phi}\right)^T \quad (7)$$

For a better discrimination the criterion should have the significant discriminant information in it. However in the above the between-class scatter matrix is computed at the expense of classes that are close to each other, leading to significant overlap between them. Thus the required modification in between-class scatter matrix is that it has to consider the separability parameter of a sample with other classes. Guangoai et al. proposed some boosting parameters which enhance the performance of scatter operators [22]. One of the parameter is, the pair wise class discriminant distribution which measures the distance between the classes in terms of the above said separability parameter. The value of this pair wise class discriminant distribution is based on the mislabel distribution of samples in the classes [21]. The pair wise class discriminant distribution $d_{i,j}$ between classes X_i and X_j can be calculated as follows:

$$d_{i,j} = \begin{cases} \frac{1}{2}\sum_{m=1}^{N_i}\Gamma\left(x_m^i, j\right) + \sum_{n=1}^{N_j}\Gamma\left(x_n^j, i\right) \\ i \neq j \\ 0, otherwise \end{cases}$$

(8)

Here the mislabel distribution $\Gamma\,(*,\bullet)$ Measures the extent of the difficulty of discriminating the example * from the improper label ●. Obviously, a larger value of $d_{i,j}$ indicates the worse separability between two classes X_i and X_j. Therefore this parameter is used to compute the between class scatter matrix $S_B^{i\phi}$. Thus the modified between-class scatter matrix $S_B^{i\phi}$ is

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

(9)

Where the weighting function w(*) is generally chosen to be monotonically increasing function of *. For simplicity, we let w(*)=* in this paper. Based on the above processing, it can be seen that classes that are not well separated in high dimensional space and thus can potentially impair the classification performance should be more heavily weighted in this space.

Another scatter matrix, the within-class scatter $S_W^{i\phi}$ computation procedure considers only the distance between the samples in that class. This will fail to estimate the correct value for improved classification, due to minimizing the spread of the outlier classes while neglecting the minimization of other co-variances. Inorderto solve this problem the parameter r_i measures the relevance based weight for class i and the other parameter q_{im} which measures the hardest samples in class i are considered in the computation procedure of within-class scatter operator $S_W^{i\phi}$. These operators are described in [22]. The modified within-class scatter operator $S_W^{i\phi}$ is defined as

$$S_W^{i\phi} = \sum_{m=1}^{N_i} r_i\, q_{im}\left(\phi\left(x_m^i\right) - \mu_i^\phi\right)\left(\phi\left(x_m^i\right) - \mu_i^\phi\right)^T$$

m=1,…..N_i, i=1,….C

(10)

Where $r_i = \sum_{j\neq i} w\left(d_{i,j}\right)$ is the relevance-based weight for class X_i and it shows the influence of class X_i in the estimation of $S_W^{i\phi}$ [25]. $q_{im} = w\left(\sum_{j\neq i}\Gamma\left(x_m^i, j\right)\right)$, indicates a greater difficulty of classifying a sample of class i. Thus the parameter q_{im} shows the difficult examples in the class i.

From the above, the parameters $d_{i,j}$, r_i , q_{im} enhance the estimation of within-class and between-class scatter operators. Thus these parameters are called boosting parameters. These modified scatter operators $S_W^{i\phi}$, $S_B^{i\phi}$ are called as boosted scatter operators. In this paper, we use these boosted scatter operators in the calculation of common vector of each class in the transformed space.

4. Proposed Improved Kernel Common Vector method: (IKCV)

Based on the weighted scatter operators $S_W^{i\phi}$, $S_B^{i\phi}$ we propose the modified IKCV algorithm called Improved KCV algorithm. In IKCV algorithm, the KCV technique with modified scatter matrices is used to extract common vector of a class in feature space F. In addition, common vector method is a strong feature extraction technique for classification. In general, some sampling procedures are employed to artificially weaken the discriminant technique and in IKCV, we choose some examples in each class based on q_{im} to focus hardest examples in each class. Then the improved KCV technique is applied on these classes for feature extraction. The kernel common vector of each class can be computed by using our proposed algorithm.

Thus IKCV method classifies a new test image x_{test} to class C by finding the minimum Euclidean distance between Ω_{test}^ϕ *and* Ω_i^ϕ *i.e.,*

$$\hat{C} = \min_i \left\|\Omega_i^\phi - \Omega_{test}^\phi\right\|, i = 1 \ldots c$$

(11)

Algorithm:-

Assuming that in a set $X = \left\{\left\{x_1^1, x_2^1, \ldots x_N^1\right\}, \left\{x_1^2, x_2^2, \ldots x_N^2\right\}, \ldots \left\{x_1^c, x_2^c, \ldots x_N^c\right\}\right\}$ 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 [27] and [28], the training samples in X are transformed into an implicit higher dimensional feature space F through a non linear mapping $\phi(x)$. Let the sample matrix be x= [x_1,x_2…………x_n] which becomes $X_\phi = \left[\phi(x_1), \phi(x_2), \ldots\ldots\phi(x_n)\right]$ after samples are mapped into feature space F through a non-linear mapping. The feature vector in F is to be computed by computing the inner product of two vectors in F with a kernel function $k(x, y) = \Phi(x)^T \Phi(y)$.

The modified within class scatter matrix $S_W^{i\phi}$, the modified between class scatter matrix $S_B^{i\phi}$ and the total scatter matrix of a class is defined by using the boosting parameters as,

$$S_W^{i\phi} = \sum_{m=1}^{N_i} r_i\, q_{im} \left(\phi(x_m^i) - \mu_i^\phi\right)\left(\phi(x_m^i) - \mu_i^\phi\right)^T \quad (12)$$

$$S_B^{i\phi} = \sum_{i=1}^{C-1} \sum_{j=I+1}^{C} \frac{N_i N_j}{N^2} w(d_{i,j})(\mu_i^\phi - \mu_j^\phi)(\mu_i^\phi - \mu_j^\phi)^T \quad (13)$$

$$S_T^{i\Phi} = S_B^{i\phi} + S_W^{i\phi} \quad (14)$$

The total scatter matrix of all classes is defined as

$$S_T^{\Phi} = \sum_{i=1}^{C} S_T^{i\Phi} \quad (15)$$

Where μ^ϕ is mean of all samples and μ_i^ϕ is the mean of samples of the i[th] class in F. The $d_{i,j}$, r_i and q_{im} are called boosting parameters which are described in previous section.

The algorithm for IKCV is summarized as follows.

STEP 1: In transformed feature space F, the training set samples of class i is trasformd on to R(S_T^Φ) using the kernel PCA.

STEP 2: Compute the new within-class scatter matrix $\widetilde{S}_W^{i\Phi}$ as

$$\widetilde{S}_W^{i\phi} = \sum_{m=1}^{N_i} r_i\, q_{im} \left(\phi(x_m^i) - \widetilde{\mu}_i^\phi\right)\left(\phi(x_m^i) - \widetilde{\mu}_i^\phi\right)^T \quad (16)$$

in the reduced space.

STEP 3: Compute the new between-class scatter matrix $\widetilde{S}_B^{i\phi}$ as

$$\widetilde{S}_B^{i\phi} = \sum_{i=1}^{C-1} \sum_{j=I+1}^{C} \frac{N_i N_j}{N^2} w(d_{i,j})(\widetilde{\mu}_i^\phi - \widetilde{\mu}_j^\phi)(\widetilde{\mu}_i^\phi - \widetilde{\mu}_j^\phi)^T \quad (17)$$

in the reduced space.

The kernel matrix $K^{(i)}$ is defined as $\Phi^T \Phi^{(i)} = (K^{(i)j})$, j=1,…,C and each submatrix $k^{(i)j}$ is defined as [18],

$k^{(i)j} = (k_{mn}^{(i)j})$, m=1,…,$N_j$, n=1,….,$N_i$

STEP 4: For each class, find a basis of the null space of $\widetilde{S}_W^{i\Phi}$. This can be done by Eigen decomposition. The normalized eigen vectors corresponding to the zero eigen values of $\widetilde{S}_W^{i\Phi}$ form an orhtonormal basis for the null space of $\widetilde{S}_W^{i\Phi}$. Let $Q^{(i)}$ be a matrix whose columns are the computed eigen vectors corresponding to the zero eigen values.

STEP 5: The matrix of basis vectors $W^{(i)}$, whose columns span the intersection subspace of the i[th] class .

The number of basis vectors spanning the intersection subspace is determined by the dimensionality of N($\widetilde{S}_W^{i\Phi}$) for each class.

STEP 6: The extracted features all classes are available in the matrix W. The feature of class I is available in $W^{(i)}$. Thus the common feature of class I is obtained by projecting any sample in that class on to its projection matrix. This is defined as

$$\Omega_{com}^{(i)} = W^{(i)^T} \Phi(x_m^i) \quad (18)$$

i=1,…..,C, m=1…N_i

Note that the above step independent of m that is the common vector of a class is independent of samples in that class.

STEP 7: To recognize a given test sample, we compute the feature vector of a test sample by

$$\Omega_{test}^{(i)} = W^{(i)^T} \Phi(x_{test}^i) \quad (19)$$

STEP 8: This method classifies a new test image x_{test} to class C by finding the minimum Euclidean distance between Ω_{test}^ϕ *and* Ω_i^ϕ *i.e.,*

$$\hat{C} = \min_i \left\| \Omega_i^\phi - \Omega_{test}^\phi \right\|, i = 1 \ldots c$$

5. Experimental Results:

The face images of YaleB university database [24] is used to test our proposed method. The sample training images are shown in fig.1. In this portion of the face database were retrieved for our training and testing purpose. The retrieved face database consists of images from C=12 different people, using 10 images from each person, for a total of 120 images. The image contains variations with the following facial expression as center-light, left-light, and normal, right-light, happy, sad 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

ISSN:0974-8652

occurred in the same co-ordinates of the image. The resulting images were then cropped. The final size of the image was 92 ×112.

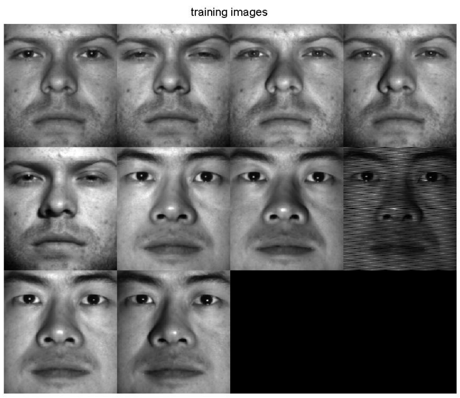

fig. 1. Sample training images from yaleB face database

The training set consists 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 84 images and a test set of 36 images were created. This recognition process was repeated 5 times and 5 different training and test sets were created. These 5 training set and testing set were created by randomly selecting samples from classes at each trial. The final recognition rates for the experiment were found by averaging these rates that were obtained in each trial. An appropriate selection of kernel functions give rise to different constructions of the implicit feature space [25] we have experimented with polynomial kernels $k(x,y)=(<x,y>)^n$ of degree n=2.

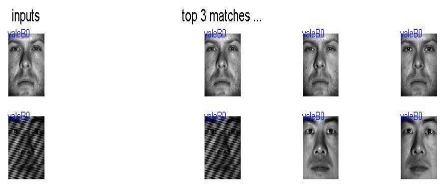

fig.2. Sample test result identified the given test image in a trial

The sample test result is shown in fig-2. The fig-2 shows that the proposed method identified the given test image varying in different lighting conditions and also varying in facial expression. The top few matches of a given test image is found and displayed along with its label. The IKCV method significantly outperformed the KCV method and Modified CV method in all cases ie., in terms of accuracy and in recognizing time.

We can apply both the Modified CV and KCV methods also here since the dimensionality of the sample space, which is 10304, is much larger than the total number of samples in the training set. The computed recognition rates on YaleB face database of KCV, Modified CV and IKCV are given in table 1:

Non-linear classifiers	Recognition rates (%)
Kernel CV	91.71
Modified CV	92.26
IKCV	94.26

Table .1. Comparison of average recognition rates (%)

In terms of classification accuracy, the Improved kernel CV method achieved the highest recognition rate than other non-linear methods. The recognition results in Table 1 show that the IKCV method outperformed the other CV methods for YaleB face database. The reason for this is that the classes were better represented by the unique subspace that were obtained using all the data samples in the training set by the IKCV method. Among subspace classifiers, improved kernel common vector method (IKCV) outperforms other methods in the case of classes having different co-variance structures in a high dimensional sample space.

6. Conclusion:

In this paper, we proposed a new algorithm for face recognition which addresses the non balanced problems in KCV. We showed that the weighted between class scatter matrix and weighted within class scatter matrix solves the problems of KCV. These matrices are calculated by using the pair wise class discriminant information which improves the classification accuracy. The hardest examples for training are selected based on the mislabel distribution is also increase the recognition performance in testing time. Our proposed algorithm gives an ensemble-based KCV framework with strong nonlinear feature extraction capability, and simultaneously overcomes the non-balanced and small sample size problems commonly encountered by KCV based methods. The experimental result shows that the proposed IKCV enhances the recognition performance of the KCV.

Acknowledgement
We would like to express our thanks to referees for valuable comments that improved the paper.

References

[1] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," Proc. IEEE, vol. 83, pp. 705-740, May 1995.
[2] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face Recognition: A Literature Survey," Technical Report CAR-TR-948, Univ. of Maryland, College Park, 2000.
[3] W. Zhao, R. Chellappa, and A Krishnaswamy, "Discriminant Analysis of Principal Components for Face Recognition," Proc.Third IEEE Int'l Conf. Automatic Face and Gesture Recognition, pp. 336-341, Apr. 1998.
[4] E.Oja, Subspace Methods of Pattern Recognition. New York:Res.Stud. Press, 1983.

[5] J.Laaksonen, " Subspace classifiers in recognition of handwritten digits", Ph.D.dissertion,Helsinki Univ.Technol., Espoo, Finland, 1997.
[6] S.-W. Kim and B.J.Oommen, "On utilizing search methods to select subspace dimensions for kernel-based nonlinear subspace classifiers", IEEE Trans. Pattern Anal. Mach. Intell., vol.27, no.1, pp.136-141, Jan. 2005.
[7] S.Watanabe, P.F. Lambert, C.A. Kulikowski, J.L.Buxton, and R.Walker, "Evaluation and selection of variables in pattern recognition", in Compter and Information Sciences ii. New York: Academic, 1967, p.91.
[8] S.Watanabe and N.pakvasa, " Subspace method in pattern recognition", in Proc. 1st Int. Conf. Pattern Recog., Washington, DC,1973,pp.25-32.
[9] K.Fukunaga and W.L.Koonz. "Application of the Karhuenn-Loeve expansion to feature selection and ordering", IEEE Trans. Comput., vol.C-19, no.4,pp.311-318,Apr.1970.
[10] M.Turk and A.P. Pentland, "Eigenfaces for recognition" j.cogn.Neurosci., vol.3, no.1, pp. 71-86,1991.
[11] R.a. Fisher, " The use of multiple measurements in taxonomic problems", Ann. Eugen., vol.7, no.2, pp. 179-188, 1936.
[12] M.B.Gulmezoglu, V.Dzhafarov, and A.Bakana , "The common vector approach and its relation to principle component analysis:, IEEE Trans. Speech Audio Process., vol9, no.6, pp.655-662, sep.2001.
[13] M.B.Gulmezoglu, V.Dzhafarov, M.keskin and A.Bakana , "A novel approach to isolated word recognition", IEEE Trans. Speech Audio Process., vol 7, no.6, pp.620-628, Nov. 1999.
[14] H.Cevikalp, B.Barkana, and A.Barkana, "A comparison of the common vector and the discriminative common vector methods for face recognition", in Proc. 9th World Multi-Conf. systemics, Cybern. and Inf., Orlando, FL, 2005.
[15] T.Balachander and R.Kothari, "Kernel based subspace pattern classification", in Proc. Int. Joint Conf. Neural Netw., 1999, vol.5, pp. 3119-3122.
[16] K.Tsuda, "Subspace classifier in reproducing kernel Hilbert space,"in Proc. Int. Joint Conf. Neural Netw., 1999, vol.5, pp.3454-3457.
[17] B.Scholkopf, A.J.Smola, and K.R. Muller, "Nonlinear component analysis as a kernel eigen value problem," Neural Comput., vol.10, no.5, pp.1299-1319, Jul.1998.
[18] Hakan Cevikalp, Marian Neamtu, and Atalay Barkana, " The Kernel Common vector method: A novel Nonliear Subspace Classifier for Pattern Recognition,"IEEE Trans. Systems, Man, and Cybernetics, vol.37, no.4, Aug.2007.
[19] Rifat Edizkan, M.B. Gulmezoglu, Semih ergin and Atalay Barkana, "Improvements on common vector approach for multi class problems", study report of the project (200315053) supported by the Research Fund of Osmangazi University.
[20] G.Baudat and F.Anouar, "Generalized discriminant analysis using a kernel approach", Neural Computation, 12:2385-2404, 2000.
[21]Y.Freund and R.E.Schapire, " A decision therotic generalization of online learning and an application to boosting", Jornal of Computer and System Sciences, 55(1):119-139, 1997.
[22] Guang Dai and Dit-Yan Yeung , "Boosting Kernel Discriminant Analysis and Its Application to Tissue Classification of Gene Expression Data", IJCAI-07, PP. 744-749.
[23] e.k.Tang, P.N.Suganathan, X. Yao, and A.K.Qin, "Linear dimensionality reduction using relevance weighted lda", Pattern Recognition, 38(4):485-493, 2005.
[24] Georghiades, A.S and Behumeur, P.N and Kriegman. D.J, "from few to many : illumination cone models for face recognition under variable lighting and pose", IEEE Trans. Pattern Analysis & machine Intelligence, vol.23(6), pp.643-660 , 2001.
[25] F.Perez-Cruz and O.Bousquet, "Kernel methods and thir potential use in signal processing", IEEE Signal Process. Mag., vol. 21, no.3, pp.57-65, May 2004.
[26] Andrea F.Abate, Michael Nappi, Daniel Riccio, Gabrido Sabatino, "2D and 3D face recognition : A Survey", Pattern Recognition letters 28(2007) 1885-1906.
[27] K-R Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms", IEEE Transactions on Neural Networks, Vol. 12, No. 2, pp. 181-201, March 2001.
[28] J. Shawe-Taylor and N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge Univ. Press, England, 2004.