

# On Using a Pre-clustering Technique to Optimize LDA-Based Classifiers for Appearance-Based Face Recognition\*

Sang-Woon Kim<sup>1</sup> and Robert P.W. Duin<sup>2</sup>

<sup>1</sup> *Senior Member, IEEE.* Dept. of Computer Science and Engineering,  
Myongji University, Yongin, 449-728 South Korea  
kimswo@mju.ac.kr

<sup>2</sup> Information and Communication Theory group,  
Delft University of Technology, The Netherlands  
r.p.w.duin@tudelft.nl

**Abstract.** Fisher's Linear Discriminant Analysis (LDA) is a traditional dimensionality reduction method that has been proven to be successful for decades. To enhance the LDA's power for high-dimensional pattern classification, such as face recognition, numerous LDA-extension approaches have been proposed in the literature. This paper proposes a new method that improves the performance of LDA-based classification by simply increasing the number of (sub)-classes through clustering a few of classes of the training set prior to the execution of LDA. This is based on the fact that the eigen space of the training set consists of the range space and the null space, and that the dimensionality of the range space increases as the number of classes increases. Therefore, when constructing the transformation matrix, through minimizing the null space, the loss of discriminative information resulted from this space can be minimized. To select the classes to be clustered, in the present paper, the *intra*set distance is employed as a criterion and the *k*-means clustering is performed to divide them. Our experimental results for an artificial data set of XOR-type samples and a well-known benchmark face database of Yale demonstrate that the classification efficiency of the proposed method could be improved.

**Keywords:** Face Recognition, Linear Discriminant Analysis (LDA), Pre-clustered LDA (PLDA).

## 1 Introduction

Over the past two decades, numerous families and avenues for Face Recognition (FR) systems have been developed. This development is motivated by the broad

---

\* The work of the first author was partially done while visiting at Delft University of Technology, 2628 CD Delft, The Netherlands. This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government(MOST) (F01-2006-000-10008-0).

range of potential applications for such identification and verification techniques. Recent surveys are found in the literature [1] and [2] related to FR. As facial images are very high-dimensional, it is necessary for FR systems to reduce these dimensions. Linear Discriminant Analysis (LDA) is one of the most popular linear projection techniques for dimension reduction [3]. LDA has been widely adopted owing to its efficiency, but it does not capture nonlinear manifolds of faces which exhibit pose and background variations<sup>1</sup>. When using LDA methods for dimensionality reduction, thus, they sometimes fail to discover the intrinsic dimension of the image space. To enhance the LDA's power for FR, numerous LDA-extensions including two-stage LDA [3], direct LDA [4], and kernel-based LDA [5] have been proposed in the literature. Beside these, to discover the nonlinear manifold structure, various techniques including LLE (Locally Linear Embedding) [6], LLDA (Locally Linear Discriminant Analysis) [10], SDA (Subclass Discriminant Analysis) [11], and Dissimilarity-Based Classifiers (DBC) [12] have been proposed<sup>2</sup>.

The approach that is proposed in this paper concerns treating the nonlinear manifolds of each face class as a set of subclasses when the class is represented by multiple separated Gaussian distributions. LDA methods effectively use the concept of a within-class scatter matrix,  $S_w$ , and a between-class scatter matrix,  $S_b$ , to maximize a separation criterion, such as  $J = \text{tr}(S_w^{-1}S_b)$ . It is well-known that for  $c$ -class (and  $d$ -dimensional) problems,  $S_b$  has the rank  $c - 1$ , and subsequently the rank of  $S_w^{-1}S_b$  is  $c - 1$  [13]. This means that there are, at most,  $c - 1$  eigenvectors corresponding to nonzero eigenvalues since the rank of the matrix  $S_b$  is bounded by  $c - 1$ . When solving an eigenvalue problem on the matrix  $S_w^{-1}S_b$ ,  $c - 1$  eigenvalues are nonzero and the others are zero. Thus, without losing the criterion value, the  $d$ -dimensional input space can be mapped onto the  $(c - 1)$ -dimensional subspace spanned by the  $c - 1$  eigenvectors corresponding to these nonzero eigenvalues. However, the discarded  $(d - c + 1)$ -dimensional subspace spanned by the  $d - c + 1$  eigenvectors, which are corresponding to these zero eigenvalues, may also contain a useful piece of information for discrimination.

From this consideration, in the present paper, a new way of improving the performance of LDA-based classification is proposed. This improvement is achieved by incorporating the clustering technique into the LDA reduction process. A few classes of a training set are first selected. Each of the selected classes is clustered into a set of subclasses by invoking a dividing (or clustering) method prior to the execution of LDA. Consequently, the number of classes of the training set increases from  $c$  to  $c'(\geq c)$ . In this approach, the major problems to be addressed are those of selecting the classes to be clustered and determining the optimal number of Gaussians per class, i.e., the number of clusters. To solve these problems, in this paper, an Euclidian distance called as *intra-set distance* and

---

<sup>1</sup> In FR, it is well-known that facial images to be recognized have many kinds of variations, such as pose, illumination, facial expression, distance, and background.

<sup>2</sup> Concerning Linear Dimensionality Reduction (LDR), there are numerous variants of Fisher's LDR. The details of the variants of LDR can be found in the well known literature including [7], [8], and [9].

a  $k$ -means clustering algorithm are employed. To select the classes, the intraset distances for all classes are first computed. After sorting them in the descending order, some of the highest variances (i.e., the largest distance) are selected and divided into subclasses by invoking the  $k$ -means clustering algorithm. Finally, a classification is performed by invoking a classifier built in the  $c'$ -class feature space, where the dimensionality is *maximally*  $c' - 1$ .

The main contribution of this paper is to demonstrate that the performance of LDA-based classifiers can be improved by employing a pre-clustering step. This has been done by incorporating the clustering technique into the LDA reduction process and by demonstrating its power in classification accuracy. The reader should observe that this philosophy is *quite* simple and distinct from those used in the recently-proposed LLE [6], LLDA [10], or SDA [11] strategies.

## 2 Optimizing LDA-Based Classification

The fundamental problem that we encounter when classifying the high-dimensional task is that of reducing the dimensionality of the training samples. One of the most popular approaches to solve this problem is employing the LDA. A conventional LDA-based classification method <sup>3</sup> is summarized in the following:

1. Compute two scatter matrices,  $S_w$  and  $S_b$ , from the training samples,  $T$ . Following this, construct a transformation matrix,  $W$ , by keeping the eigenvectors corresponding to the nonzero eigenvalues of  $\text{tr}(S_w^{-1}S_b)$ .
2. Project the training set  $T$  into a reduced dimensional feature space by using the transformation matrix  $W$ , where the dimensionality of the feature space is *maximally*  $c - 1$ . To test a sample  $\mathbf{z}$ , compute a feature vector,  $\mathbf{z}'$ , using the same transformation matrix.
3. Achieve a classification based on invoking a classifier built in the feature space and operating on the vector of  $\mathbf{z}'$ .

In the above algorithm, the classification performance can be improved by increasing the number of classes from  $c$  to  $c'$ . This increase can be obtained by dividing the manifolds of each class into a set of clusters. Rather than divide *all* classes of  $T$ , a few of the classes are bisected in this paper. To choose the classes to be clustered, the distribution variance of each class can be used as a criterion. To measure the variance of the class, the so-called intra-set distance is used. The distance, which is an Euclidean distance, is briefly introduced here.

**Intraset Distance:** Let  $T = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathbb{R}^d$  be a set of  $n$  feature vectors in a  $d$ -dimensional space. Assume that  $T$  is a labeled data set so that  $T$  can be decomposed into, for example,  $c$  disjoint subsets  $\{T_1, \dots, T_c\}$  such that  $T = \bigcup_{k=1}^c T_k, T_i \cap T_j = \phi, \forall i \neq j$ . Then, a criterion associated with  $T_i$  is defined

<sup>3</sup> For want of a better term, LDA-based classification performed without utilizing the pre-clustering technique will be referred to as “conventional” schemes.

as follow: For an arbitrary pattern  $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jd})^T \in T_i$ , the mean of  $d(\mathbf{x}_j, T_i - \{\mathbf{x}_j\})$  over  $T_i$  is called the *intra-set distance* of  $T_i$  and is denoted by

$$\begin{aligned}
 D^2(T_i) &= \frac{1}{n_i} \sum_{j=1}^{n_i} d^2(\mathbf{x}_j, T_i - \{\mathbf{x}_j\}) \\
 &= \frac{1}{n_i(n_i - 1)} \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \sum_{k=1}^d (x_{jk} - x_{lk})^2.
 \end{aligned}
 \tag{1}$$

By conveniently rearranging the elements in the triple summation of (1) and considering the relations of  $\overline{x_{jk}} = \overline{x_{lk}}$  and  $\overline{(x_{jk})^2} = \overline{(x_{lk})^2}$  for arbitrary  $j$  and  $l$ , the intra-set distance can be expressed in terms of the unbiased variances of components of the given patterns like:  $D^2(T_i) = 2 \sum_{k=1}^d \sigma_k^2$ , where  $\sigma_k^2 = \frac{n_i}{n_i - 1} \left\{ \overline{(x_{jk})^2} - (\overline{x_{jk}})^2 \right\}$  for all  $\mathbf{x}_j \in T_i$ . Details of the derivation are omitted here in the interest of compactness, but can be found in [14]. This is the rationale of the scheme for employing the intra-set distance as a criterion to select the classes to be clustered. To simplify the classification task, in this paper, the chosen classes are divided into *two* clusters. However, this can be expanded into a general solution by utilizing a way of automatically determining the number of clusters and measuring the goodness of the clusters formed. The details of this subject are omitted here, but can be found in [15].

**Pre-clustered LDA-Based Classification:** As mentioned previously, by the means of clustering the nonlinear manifolds of a class into a set of subclasses, the classification performance of the above conventional method can be improved furthermore. To achieve this, first of all, the intra-set distances for all classes are computed to choose the class to be clustered. The proposed approach, which is referred to as a Pre-clustered LDA-based (PLDA) classification, is summarized in the following:

1. Compute the intra-set distances of the input training data set  $T_i$  for all  $i$ ,  $1 \leq i \leq c$ , and sort them with the intra-set distances in descending order,  $D^2(T_1) \geq D^2(T_2) \geq \dots \geq D^2(T_c)$ . Then, set  $c' = c$  and  $\theta_1 = 1$ .
2. Do the following : Increasing  $i$  by unity from 1 to the number of classes in  $T$  per epoch:
  - (a) Cluster the training set of class  $\omega_i$ ,  $T_i$ , into two clusters, which are named as  $T_i$  and  $T_{i+c}$ , and increase the number of *sub*-classes,  $c' = c' + 1$ .
  - (b) Compute two scatter matrices,  $S_w$  and  $S_b$ , from the training samples,  $T = T_1 \cup T_2 \cup \dots \cup T_{i+c}$ . Following this, construct a transformation matrix,  $W$ , by keeping the eigenvectors corresponding to the nonzero eigenvalues of  $tr(S_w^{-1} S_b)^4$ .

---

<sup>4</sup> This step can be different depending on which method is employed. For example, in the direct LDA, this step is performed after doing the *simultaneous* diagonalization of  $S_b$  and  $S_w$  [13].

- (c) Project the training set  $T$  into a reduced dimensional feature space by using the transformation matrix  $W$ , where the dimensionality of the feature space is *maximally*  $c' - 1$ . To test a sample  $z$ , compute a feature vector,  $z'$ , using the same transformation.
  - (d) Achieve a classification based on invoking a classifier built in the feature space and operating on the vector of  $z'$ . The classification accuracy for all testing samples is labeled as  $\theta_2$ . If  $\theta_1 - \theta_2 > \rho$ , then go to the next step. Otherwise, Step 2 is repeated after  $\theta_1 = \theta_2$ .
3. Output  $\theta_2$  as the final result, then the process terminates.

In Step 2 of the above algorithm, to simplify the classification task for the paper, the maximum iteration number can be limited experimentally. Also, in the same step, the threshold value  $\rho$  is determined experimentally.

The computational complexity of the proposed algorithm depends on the computational costs associated with the number of training samples and the clustering algorithm. The time complexity of PLDA can be analyzed as follows: Step 1 requires  $O(n^2 \times d) + O(c)$  time. Step 2 requires  $O(n) + O(8d^3) + O(n \times d) + O(\gamma_1)$  time (where  $\gamma_1$  is the time for training a classifier with the training set  $T$ ) to perform the four sub-steps of computing transformation matrix, projecting data samples onto divided sub-class space, and classifying a test sample with the classifier designed in the sub-class space. Step 3 requires  $O(1)$  time to terminate the algorithm. Thus, the total time complexity of the PLDA is  $O(n^2d + d^3 + \gamma_1)$ . Then, the space complexity of PLDA is  $O(nd)$ .

### 3 Experimental Results : Artificial/Real-Life Data Sets

The proposed method has been tested and compared with conventional methods. This was done by performing experiments on an artificial data set (which is named as XOR4) and a well-known face database, namely, the Yale<sup>5</sup> database.

The data set named “XOR4”, which has been included in the experiments as a baseline data set, was generated from a mixture of four 4-dimensional Gaussian distributions as follows: (1)  $p_1(x) = \frac{1}{2}N(\mu_{11}, I_4) + \frac{1}{2}N(\mu_{12}, I_4)$  and (2)  $p_2(x) = \frac{1}{2}N(\mu_{21}, I_4) + \frac{1}{2}N(\mu_{22}, I_4)$ , where  $\mu_{11} = [-2, -2, 0, 0]$ ,  $\mu_{12} = [2, 2, 0, 0]$ ,  $\mu_{21} = [2, -2, 0, 0]$ , and  $\mu_{22} = [-2, 2, 0, 0]$ . Also,  $I_4$  is the 4-dimensional *Identity* matrix. Here, it is clear that each class contains *two* clusters. Thus, this case is better treated as a *four*-class problem rather than a *two*-class one.

The face database captioned “Yale” contains 165 gray scale images of 15 individuals. There are 11 images per subject, one of each for the following facial expression or configurations: center-light, left-light, right-light, with glasses, without glasses, happy, normal, sad, sleepy, surprised, and winking. The size of each image is  $243 \times 320$  pixels for a total dimensionality of 77760. In this experiment, to reduce the computational complexity, each facial image of Yale database was down-sampled into  $61 \times 80$  and then represented by a centered vector of normalized intensity values. To obtain a different data set, a part of the

<sup>5</sup> <http://www1.cs.columbia.edu/belhumeur/pub/images/yalefaces>

down-sampled image, which is  $28 \times 23$  size and contains only facial components without background, was extracted. This face database is termed as “Yale2” in a subsequent section.

**Experimental Method:** In this paper, all experiments were performed using a “leave-one-out” strategy. To classify an image of object, that image was removed from the training set and a transformation matrix,  $W$ , was computed with the  $n - 1$  images. Following this, all of the  $n$  images in the training set and the test object were translated into a  $(c' - 1)$ -dimensional space using the transformation matrix and recognition was performed based on the algorithm in Section 2. After repeating this  $n$  times for every sample, a final result was obtained by averaging them.

To perform the transformation and classification, first of all, the intraset distances for all classes were computed. After computing the distances, the data sets of the classes were sorted in the decreasing order using the intraset distance. Then, the classes of the largest distances were selected and divided into sub-classes by invoking a  $k$ -means clustering algorithm. Finally, a classification was obtained by designing a classifier in the sub-classes feature space. In this experiment, to simplify the classification task for the paper, only two, three, or four classes were selected and each of them was divided into *two* sub-classes<sup>6</sup>. Furthermore, to reduce the dimensionality of the clustered data samples, only one approach, namely, the direct LDA, was implemented. However, other approaches including PCA+LDA [3], LDA/GSVD [16], and LDA/QR [17], Kernel-based LDA [5] could have also been considered.

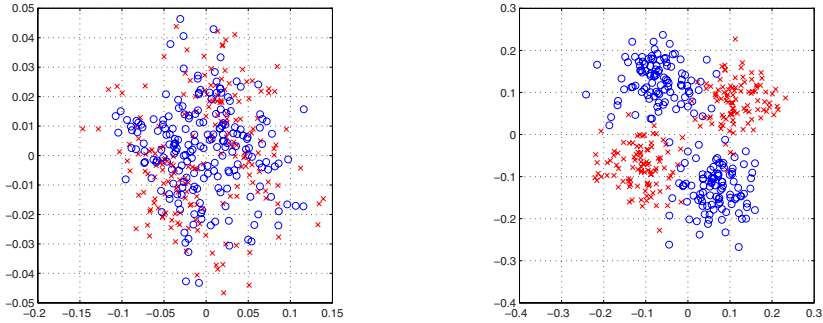
On the other hand, to maintain the diversity between the LDA-based classifications, different classifiers, such as the  $k$ -Nearest Neighbor Classifiers ( $k = 1, 3$ ), Nearest Mean Classifiers, Normal Density based Linear Classifier, Uncorrelated Normal Density based Quadratic Classifier, Parzen Density based Classifier, Fisher’s Least Square Linear Classifier, and Neural Network Classifier were designed. These classifiers are implemented with PRTools<sup>7</sup> and will be denoted as 1-NN, 3-NN, NMC, LDC, UDC, PZC, FSC, and NNC, respectively, in a subsequent section. Here, NNC is a feed-forward neural net classifier with one hidden layer of 20 sigmoid neurons.

**Experimental Results:** The run-time characteristics of the proposed algorithm for the artificial data set, XOR4, and the benchmark database, Yale, are reported below and shown in Table 1. The rationality of employing a pre-clustering step in the LDA-based classifiers is investigated first. Following this, a comparison is made between the conventional LDA-based method and the proposed scheme.

First, the results of the dimensionality reduction obtained in Step 2(c) of the proposed LDA-based scheme in Section 2 were probed into. Fig. 1 shows plots of two 2-dimensional data sets obtained from the original 4-dimensional data set

<sup>6</sup> In this experiment, to implement the normal-density based classifiers, the samples of at least two points were compulsorily assigned to each sub-class.

<sup>7</sup> PRTools is a Matlab Toolbox for Pattern Recognition. PRTools can be downloaded from the PRTools website, <http://www.prtools.org/>



**Fig. 1.** Plots of the 2-dimensional data set obtained from the original 4-dimensional artificial data set, namely, XOR4. In both pictures, the left is reduced with the direct LDA and the right is obtained with the proposed method, in which the LDA is carried out after doing a pre-clustering.

of XOR4. Here, the left is of the conventional scheme and the right is of the proposed method, respectively. In the conventional scheme, the direct LDA was applied to the original 4-dimensional data set, while the direct LDA was carried out after doing a pre-clustering for the data set in the proposed method.

From the figure, it should be observed that the accuracy of the dimensionality reduction step for the artificial data set can be improved by employing the philosophy of a pre-clustering. This is clearly shown in the classification boundary built between the two classes represented with “x” and “o” in both pictures. This characteristic could be also observed from the real benchmark face database. For the fifteen classes of Yale, the intraset distances obtained with Eq. (1) are sorted as follow:

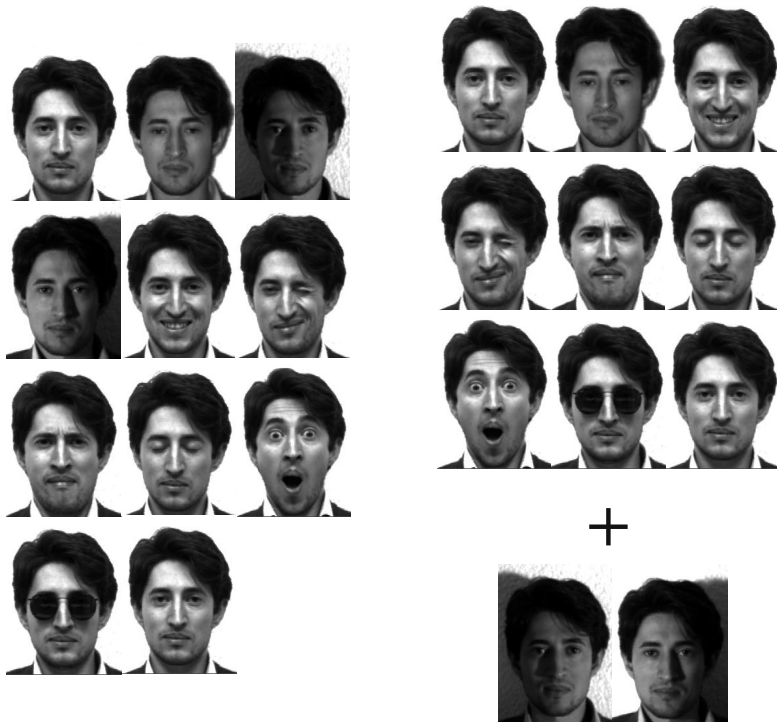
```

1.0e+007 *
2.9493  2.9240  2.9036  2.8759  2.5882  2.5647  2.1206  2.0010
1.9686  1.9128  1.7468  1.6376  1.6368  1.5628  1.2592
    
```

From these figures, some of the classes which are ranked among the largest distances are selected and clustered into two sub-classes. Fig. 2 shows faces of an object of Yale and their clustering result. Here, the left picture shows the eleven faces of the object which has the largest distance (i.e.,  $1.0e+007 * 2.9493$ ) and the right one shows that the faces are divided into two sub-classes. Faces of the two sub-classes are different from each other in the background.

From the above consideration, the rationale of the LDA-based classification for employing the pre-clustering technique is proven to be valid.

Secondly, as the main results, it should be noted that it is possible to improve the classification performance by employing the philosophy of a pre-clustering prior to the execution of the dimensionality reduction step. Table 1 shows a comparison of the classification performances for XOR4 data set and Yale



**Fig. 2.** Faces of an individual selected from Yale database. In the picture, the left eleven faces are of the class that has the largest intraset distance, while the right picture shows that the eleven faces are divided into two sub-classes. The details of the figure are discussed in the text.

database<sup>8</sup>. Here, the abbreviations NON, LDA, and PLDA, which are, respectively, *NON*-reduction classification, the conventional *LDA*-based classification, and the *Pre-clustered LDA*-based classification, indicate the classification methods employed in this experiment. Especially, *PLDA*(2), *PLDA*(3), and *PLDA*(4) are abbreviations of the classifications performed after pre-clustering the selected two, three, and four classes, respectively. For these classification methods, the best results obtained are printed in bold.

From Table 1, it is clear that the classification accuracies for the experimental databases can be improved by employing the pre-clustering technique (see the

<sup>8</sup> In Table 1, the symbol “-” indicates that there are *no* results available due to the characteristics of the data set and the out-of-memory problem. First of all, in the implementation of the direct LDA for XOR4, we have a problem in diagonalization of  $S_b = \sum_i n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$  because two means, namely, the class mean,  $\mathbf{m}_i$ , and the global mean,  $\mathbf{m}$ , are (almost) *zero*. Then, to implement the linear/quadratic normal density based classifiers (LDC and QDC) for very high dimensional data, we need three big dimensional matrices, for example,  $10304 \times 10304$ . To be consistent with other methods, however, we did not fix the implementation.



**Table 1.** A comparison of the classification performances of LDA-based classifiers for XOR4 and Yale. Here, in the second column, the abbreviations NON, LDA, and PLDA(·) indicate *NON*-reduction classification, the conventional *LDA*-based classification, and the *Pre-clustered LDA*-based classification, respectively. Also the symbol “\_” indicates that there are *no* results available due to the characteristics of the data set. The details of the table are discussed in the text.

Dataset Names	Analysis Methods	Experimental Classifiers							
		1-NN	3-NN	NMC	LDC	UDC	PZC	FSC	NNC
XOR4	NON	89.25	92.50	46.50	45.25	44.75	93.00	45.25	85.00
	LDA	—	—	—	—	—	—	—	—
	PLDA(2)	<b>90.50</b>	<b>94.00</b>	<b>95.00</b>	<b>94.25</b>	95.25	<b>94.75</b>	<b>94.50</b>	<b>94.75</b>
	PLDA(3)	<b>90.50</b>	<b>94.00</b>	<b>95.00</b>	<b>94.25</b>	<b>95.75</b>	<b>94.75</b>	<b>94.50</b>	94.50
	PLDA(4)	<b>90.50</b>	<b>94.00</b>	<b>95.00</b>	<b>94.25</b>	<b>95.75</b>	<b>94.75</b>	<b>94.50</b>	94.00
Yale	NON	79.39	78.18	78.18	—	—	6.67	<b>95.76</b>	58.79
	LDA	89.70	90.91	92.12	92.12	6.67	4.85	81.82	83.03
	PLDA(2)	90.91	93.33	<b>94.55</b>	93.94	<b>87.88</b>	92.73	91.52	86.06
	PLDA(3)	92.73	<b>95.76</b>	93.94	<b>94.55</b>	84.85	93.94	92.73	<b>89.70</b>
	PLDA(4)	<b>93.33</b>	<b>95.76</b>	<b>94.55</b>	<b>94.55</b>	86.67	<b>94.55</b>	93.94	87.28
Yale2	NON	78.18	76.97	78.18	<b>95.15</b>	73.94	6.67	<b>89.70</b>	58.79
	LDA	90.30	89.09	92.12	92.12	<b>86.06</b>	90.30	84.24	80.00
	PLDA(2)	90.91	<b>91.52</b>	92.12	92.12	84.24	<b>92.73</b>	84.24	<b>81.82</b>
	PLDA(3)	89.70	89.70	<b>93.33</b>	92.73	83.64	90.91	85.45	79.39
	PLDA(4)	<b>92.12</b>	<b>91.52</b>	<b>93.33</b>	93.94	81.82	92.12	87.27	78.18

bold-faced ones). An example of this is the classification accuracy rates (%) obtained with 1-NN classifiers designed for Yale. The classification accuracies of NON, LDA, PLDA(2, 3, 4) methods are 79.39, 89.70, 90.91, 92.73, and 93.33 (%), respectively. Additionally, the classification accuracies of NNC classifiers are 58.79, 83.03, 86.06, 89.70, and 87.28 (%), again respectively. From this consideration, the reader can observe that the classification performance of NON was improved with LDA and further improved with PLDA. The details are omitted here in the interest of compactness. From the table, however, it should be mentioned that the classification accuracies of PLDA(2), PLDA(3), and PLDA(4) are almost the same while the number of classes to be clustered increases. For instance, the accuracy rates obtained with the 1-NN classifier for XOR4 are the same as 90.50 (%). For Yale database, the same characteristics can be observed.

From the above consideration, it should be also observed that the proposed scheme of employing the pre-clustering technique is useful in improving the classification performance.

## 4 Conclusions

In this paper, a method that seeks to improve the classification performance of LDA-based approaches for appearance-based face recognition is considered. This method involves a pre-clustering step prior to the execution of Fisher’s

criterion to find out the nonlinear structure of faces and divide it into a set of clusters. The experimental results for an artificial data set and a well-known face database demonstrate that the proposed scheme is better than the conventional ones in terms of the classification accuracy rates. Even though an investigation has been made, focusing on the possibility of the pre-clustering technique being used to solve the nonlinear-manifolds problem of FR, many problems remain. This classification performance could be further improved by developing an optimal division (or selection) method and by designing suitable classifiers in the divided-class space. The research concerning this is a future aim of the authors.

## References

1. Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.J.: Face recognition: a literature survey. *ACM Compt. Surveys* 35(4), 399–458 (2003)
2. Ruiz-del-Solar, J., Navarrete, P.: Eigenspace-based face recognition: a comparative study of different approaches. *IEEE Trans. Systems, Man, and Cybernetics-Part C* 35(3), 315–325 (2005)
3. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. and Machine Intell.* 19(7), 711–720 (1997)
4. Yu, H., Yang, J.: A direct LDA algorithm for high-dimensional data - with application to face recognition. *Pattern Recognition* 34, 2067–2070 (2001)
5. Yang, M.-H.: Kernel eigenfaces vs. kernel Fisherfaces: Face recognition using kernel methods. In: *Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 215–220 (2002)
6. Roweis, S., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500), 2323–2326 (2000)
7. Lotlikar, R., Kothari, R.: Adaptive linear dimensionality reduction for classification. *Pattern Recognition* 33, 185–194 (2000)
8. Loog, M., Duin, P.W.: Linear dimensionality reduction via a heteroscedastic extension of LDA: The Chernoff Criterion. *IEEE Trans. Pattern Anal. and Machine Intell.* 26(6), 732–739 (2004)
9. Rueda, L., Herrera, M.: A new approach to multi-class linear dimensionality reduction. In: *Martínez-Trinidad, J.F., Carrasco Ochoa, J.A., Kittler, J. (eds.) CIARP 2006. LNCS, vol. 4225, pp. 634–643. Springer, Heidelberg (2006)*
10. Kim, T.-K., Kittler, J.: Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *Pattern Anal. and Machine Intell.* 27(3), 318–327 (2005)
11. Zhu, M., Martínez, A.M.: Subclass discriminant analysis. *IEEE Trans. Pattern Anal. and Machine Intell.* 28(8), 1274–1286 (2006)
12. Kim, S.-W.: On solving the small sample size problem using a dissimilarity representation for face recognition. In: *Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2006. LNCS, vol. 4179, pp. 1174–1185. Springer, Heidelberg (2006)*
13. Fukunaga, K.: *Introduction to Statistical Pattern Recognition*. Academic Press, San Diego (1990)
14. Friedman, M., Kandel, A.: *Introduction to Pattern Recognition - Statistical, Structural, Neural and Fuzzy Logic Approaches*. World Scientific, New Jersey (1999)

15. Maulik, U., Bandyopadhyay, S.: Performance evaluation of some clustering algorithms and validity indices. *IEEE Trans. Pattern Anal. and Machine Intell.* 24(12), 1650–1654 (2002)
16. Howland, P., Wang, J., Park, H.: Solving the small sample size problem in face recognition using generalized discriminant analysis. *Pattern Recognition* 39, 277–287 (2006)
17. Ye, J., Li, Q.: A two-stage linear discriminant analysis via QR-decomposition. *IEEE Trans. Pattern Anal. and Machine Intell.* 27(6), 929–941 (2005)