On Combining Dissimilarity-Based Classifiers to Solve the Small Sample Size Problem for Appearance-Based Face Recognition*

Sang-Woon Kim[1] and Robert P.W. Duin[2]

[1] *Senior Member, IEEE.* Dept. of Computer Science and Engineering, Myongji University, Yongin, 449-728 South Korea
kimsw@mju.ac.kr

[2] Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, The Netherlands
r.p.w.duin@tudelft.nl

Abstract. For high-dimensional classification tasks, such as face recognition, the number of samples is smaller than the dimensionality of the samples. In such cases, a problem encountered in Linear Discriminant Analysis-based (LDA) methods for dimension reduction is what is known as the Small Sample Size (SSS) problem. A number of LDA-extension approaches that attempt to solve the SSS problem have been proposed in the literature. Recently, a different way of employing a dissimilarity representation method was proposed [18], where an object was represented based on the dissimilarity measures among representatives extracted from training samples instead of the feature vector itself. Apart from utilizing the dissimilarity representation, in this paper, a new way of employing a fusion technique in representing features as well as in designing classifiers is proposed in order to increase the classification accuracy. The proposed scheme is completely different from the conventional ones in terms of the computation of the transformation matrix as well as the selection of the number of dimensions. The present experimental results demonstrate that the proposed combining mechanism works well and achieves further improved efficiency compared with the LDA-extension approaches for well-known face databases involving AT&T and Yale databases. The results especially demonstrate that the highest accuracy rates are achieved when the combined representation is classified with the trained combiners.

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 range of potential applications for such identification and verification techniques.

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

Z. Kobti and D. Wu (Eds.): Canadian AI 2007, LNAI 4509, pp. 110–121, 2007.

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]. The major limitation when applying LDA is that it may encounter what is known as the Small Sample Size (SSS) problem [4], [5]. This problem arises whenever the number of samples is smaller than the dimensionality of the samples. Under these circumstances, the sample scatter matrix can become singular, and the execution of LDA may encounter computation difficulties.

In order to address the SSS issue, numerous methods have been proposed in the literature. One popular approach that addresses the SSS problem is to introduce a Principal Component Analysis (PCA) step to remove the null space of the between- and within-class scatter matrices before invoking the LDA execution. However, recent research reveals that the discarded null space may contain the most significant discriminatory information. Moreover, other solutions that use the null space can also have problems. Due to insufficient training samples, it is very difficult to identify the true null eigenvalues. Since the development of the PCA+LDA [3], other methods have been proposed successively, such as the pseudo-inverse LDA [6], the regularized LDA [7], the direct LDA [8], the LDA/GSVD [5] and the LDA/QR [9]. In addition to these methods, the Discriminative Common Vector (DCV) technique [10], has recently been reported to be an extremely effective approach to dimension reduction problems. The details of these LDA-extension methods are omitted here as they are not directly related to the premise of the present work.

Recently, a new paradigm to pattern classification has been proposed [11] - [13] based on the idea that if "similar" objects can be grouped together to form a class, the "class" is nothing more than a set of these similar objects. This methodology is a way of defining classifiers between the classes. It is not based on the feature measurements of the individual patterns, but rather on a suitable dissimilarity measure between them. The advantage of this is clear: As it does not operate on the class-conditional distributions, the accuracy can exceed the Bayes' error bound. Another salient advantage of such a paradigm is that it does not have to confront the problems associated with feature spaces such as the "curse of dimensionality", and the issue of estimating large numbers of parameters. Particularly, by selecting a set of prototypes or support vectors, the problem of dimension reduction can be *drastically* simplified.

On the other hand, combination systems which fuse "pieces" of information have received considerable attention because of its potential to improve the performance of individual systems. Various fusion strategies have been proposed in the literature and workshops[1] - excellent studies are found in [14], [15], and [16]. The applications of these systems are many. For example, consider a design problem involving pattern classifiers. The basic strategy used in fusion is to solve the classification problem by designing a *set* of classifiers, and then combining the individual results obtained from these classifiers in some way to achieve

[1] http://www.diee.unica.it/mcs/home.html

reduced classification error rates. Therefore, the choice of an appropriate fusion method can further improve on the performance of the individual method. Various classifier fusion strategies have been proposed in the literature. The decision rules commonly used are *Product*, *Sum*, *Max*, *Min*, *Median*, and *Majority vote* rules. Their details can be found in [14] and [15].

Motivated by the methods mentioned above, a combined dissimilarity-based scheme is investigated to solve the SSS problem in FR.

Recently, Kim [18] experimented the utilization of the dissimilarity representation as a method for solving the SSS problem. Apart from utilizing the dissimilarity representation, in this paper, a new way of employing a fusion technique in representing features as well as in designing classifiers is proposed. The combined dissimilarity-based scheme is completely different from the conventional ones in terms of the computation of the transformation matrix and the selection of the number of dimensions. A problem that is encountered in this paper concerns solving the SSS problem when the number of available facial images per subject is insufficient. For this reason, *all* samples are initially represented with different dissimilarity measures[2] among the samples instead of the feature vectors themselves. However, in facial images there are many kinds of variations, such as pose, illumination, facial expression, and distance. To overcome this problem, an object is classified with a combined classifier designed in the dissimilarity space.

In some cases, newly generated features based on a certain feature combination could be more informative compared to the original features. To obtain more powerful representation, in this paper, the dissimilarity representations are first combined into new ones by building an extended matrix or by simply averaging them. Then, the object is classified by invoking a group of dissimilarity-based classifiers as the *base* classifiers designed in the newly created dissimilarity space. The final decision is obtained with a *fixed* or *trained* combiner which is applied to the outputs of the base classifiers. The details of these classifiers are included in the present paper. The present experimental results for well-known face databases demonstrate that the proposed combining mechanism works well and achieves further improved efficiency results compared with the conventional LDA-extension approaches.

Two modest contributions are claimed in this paper by the authors:

1. This paper lists the first reported results that reduce the dimensionality and solve the SSS problem by resorting to the combined dissimilarity-based classifiers. Although the result presented is only for a case when the task is face recognition, the proposed approach can also apply to other high-dimensional tasks, such as information retrieval and bioinformatics.
2. The paper contains a formal algorithm in which, to improve classification performances for high-dimensional tasks, a fusion strategy in representing features as well as in designing classifiers is employed. The paper also

[2] Here, dissimilarity representations are measured with Euclidean-based metrics, such as the Euclidean distance and the regional distance, with the intent of simplifying the problem. The details of these measures will be included in the present paper.

provides experimental results by which the rationale of the dissimilarity-based scheme for employing the fusion technique is proven to be valid.

To the best of the authors' knowledge, all of these contributions are novel to a field of high-dimensional classification such as image recognition. This paper is organized as follows: An overview is initially presented of the dissimilarity representation in Section 2. Following this, the algorithm that solves the SSS problem by incorporating the use of dissimilarity representation and a fusion strategy is presented. Experimental results for the real-life benchmark data sets are provided in Section 3, and the paper is concluded in Section 4.

2 Combining Dissimilarity-Based Classifiers (DBCs)

2.1 Foundations of DBCs

Let $T = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\} \in \Re^p$ be a set of n feature vectors in a p-dimensional space. Assume that T is a labeled data set so that T can be decomposed into, for example, c disjoint subsets $\{T_1, \cdots, T_c\}$ such that $T = \bigcup_{k=1}^{c} T_k, T_i \cap T_j = \phi, \forall i \neq j$. The goal is to design a DBC in an appropriate dissimilarity space constructed with this *training data* set and to classify an input sample $\boldsymbol{z}$ into an appropriate class. To achieve this, first of all, a prototype set of class ω_i, $Y_i = \left\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{m_i}\right\}, m = \sum_{i=1}^{c} m_i$, is extracted from the training data, T_i.

Every DBC assumes the use of a dissimilarity measure, d, computed from the samples, where $d(\boldsymbol{x}_i, \boldsymbol{y}_j)$ represents the dissimilarity between two samples, $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$. The dissimilarity computed between T and Y leads to a $n \times m$ matrix, $D(T, Y)$, where $\boldsymbol{x}_i \in T$ and $\boldsymbol{y}_j \in Y$. Consequently, an object $\boldsymbol{x}_i$ is represented as a column vector as following:

$$\left(d(\boldsymbol{x}_i, \boldsymbol{y}_1), d(\boldsymbol{x}_i, \boldsymbol{y}_2), \cdots, d(\boldsymbol{x}_i, \boldsymbol{y}_m)\right)^T, 1 \leq i \leq n. \tag{1}$$

Here, the dissimilarity matrix $D(\cdot, \cdot)$ is defined as a *dissimilarity space* on which the p-dimensional object, $\boldsymbol{x}$, given in the feature space, is represented as an m-dimensional vector $d(\boldsymbol{x}, Y)$, where if $\boldsymbol{x} = \boldsymbol{x}_i$, $d(\boldsymbol{x}_i, Y)$ is the i^{th} row of D matrix. In this paper, the column vector $d(\boldsymbol{x}, Y)$ is simply denoted by $d(\boldsymbol{x})$, where the latter is an m-dimensional vector, while $\boldsymbol{x}$ is p-dimensional.

From this perspective, it becomes clear that the dissimilarity representation can be considered as a *mapping* by which $\boldsymbol{x}$ is translated into $d(\boldsymbol{x})$; thus, m is selected as sufficiently small ($m << p$), what is being worked in is essentially a space with much smaller dimensions. Based on this consideration, the mapping could be considered as a way of solving the SSS problem.

Two factors to consider for a dissimilarity representation are to select a prototype subset from the training samples and to quantify the dissimilarity between two vectors. To do these things, various representative selection methods and dissimilarity measures have been proposed in [12], [13], and [17]. The details of these are omitted here in the interest of compactness.

2.2 Classifier Fusion Strategies (CFSs)

Recently, classifier combination ("Fusion") has received considerable attention because of its potential to improve the performance of classification systems. The basic idea is to solve each classification problem by designing a *set* of classifiers, and then combining the classifiers in some way to achieve reduced classification error rates. Therefore a choice of an appropriate fusion method can further improve on the performance of the combination. Various CFSs have been proposed in the literature - excellent studies are found in [14], [15], and [16]. The CFS's decision rules of [15] are summarized here briefly.

Consider a pattern recognition problem where pattern z is to be assigned to one of the c possible classes, $\omega_1, \cdots, \omega_c$. Assume that there are M classifiers each representing the given pattern by a distinct measurement vector. Denote the measurement vector used by the ith classifier by $\boldsymbol{x}_i, i = 1, \cdots, M$. In this case, the Bayesian decision rule computes the *a posteriori* probability $p(\omega_k|\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M)$ using the Bayes theorem as follow:

$$p(\omega_k|\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M) = \frac{p(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M|\omega_k)P(\omega_k)}{\sum_{j=1}^{c} p(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M|\omega_j)P(\omega_j)}. \tag{2}$$

Let us assume that the representations used are statistically independent. Then the joint probability distribution of the measurements extracted by the classifiers can be rewritten as follow:

$$p(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M|\omega_k)P(\omega_k) = \prod_{i=1}^{M} p(\boldsymbol{x}_i|\omega_k), \tag{3}$$

where $p(\boldsymbol{x}_i|\omega_k)$ is the measurement process model of the ith representation.

Based on (2) and (3), the commonly used decision rules, such as *Product*, *Sum*, *Max*, *Min*, *Median*, and *Majority vote* rules, are obtained. Their details can be found in [14] and [15]. Although all of them can be used in a CFS, a rule used in the present experiment, namely, the *Majority vote* rule which operates under the assumption of equal priors, can be described as follows:

$$\sum_{i=1}^{M} \Delta_{ji} = \max_{1 \le k \le c} \left\{ \sum_{i=1}^{M} \Delta_{ki} \right\} \Rightarrow \boldsymbol{z} \in \omega_j, \tag{4}$$

$$\Delta_{ki} = \begin{cases} 1, \text{ if } p(\omega_k|\boldsymbol{x}_i) = \max_{1 \le j \le c} \{p(\omega_j|\boldsymbol{x}_i)\}. \\ 0, \text{ otherwise.} \end{cases} \tag{5}$$

Here, for each class ω_k, the sum of Δ_{ji} simply counts the votes received for this result from the individual classifiers. Thus the class which receives the largest number of votes is then selected as the majority decision.

The above combination schemes can be applied for combining a *set* of distinct features as well as different classifiers. Here, it is interesting to note that a number of distinct dissimilarity representations can be combined into a new one to obtain a more powerful representation in the discrimination. The idea

of this *feature* combination is derived from the possibility that discriminative properties of different representations can be enhanced by a proper fusion [12]. There are several schemes for combining multiple representations to solve a given classification problem. Some of them are : *Average*, *Product*, *Min*, and *Max* rules. The details of these methods are omitted here, but can be found in [12].

The reasons for combining several distinct dissimilarity representations and different dissimilarity-based classifiers will be exhaustively investigated in the present paper.

2.3 Combined Dissimilarity-Based Classifiers (CDBCs)

In this section, a dissimilarity-based method of classifying the high-dimensional samples without encountering the SSS problem is proposed. A simple Dissimilarity-Based Classifier (DBC) [17] consists of the following steps:

1. Select the representative set, Y, from the training set T by resorting to one of the prototype selection methods as described in [13], [17].
2. Compute the dissimilarity matrix, $D(T, Y)$, with T and Y, in which each individual dissimilarity is computed using one of the measures. To test a sample $\boldsymbol{z}$, compute a dissimilarity column vector, $d(\boldsymbol{z})$, using the same measure.
3. Achieve a classification based on invoking a classifier built in the dissimilarity space and operating on the dissimilarity vector $d(\boldsymbol{z})$.

However, in facial images there are many kinds of variations based on such factors as pose, illumination, facial expression, and distance. Thus, by simply measuring the differences of facial images for each class, it is not possible to obtain a good representation. To overcome this limitation, a classifier fusion strategy is employed. The basic strategy used in fusion is to solve the classification problem by designing a set of classifiers, and then to combine the individual results obtained from these classifiers in some way to achieve reduced classification error rates. The tangible rationale for this fusion strategy will be presented in a later section together with the experimental results.

The proposed approach, which is referred to as a Combined Dissimilarity-Based Classifier (CDBC), is summarized in the following:

1. Select the input training data set T as a representative subset Y.[3]
2. Compute dissimilarity matrices, $D^{(1)}(T, Y)$, $D^{(2)}(T, Y)$, $\cdots$, $D^{(k)}(T, Y)$, by using the k different dissimilarity measures for all $\boldsymbol{x} \in T$ and $\boldsymbol{y} \in Y$.
3. To obtain more powerful representation, combine the dissimilarity matrices, $\{D^{(i)}(T, Y)\}_{i=1}^{k}$, into new ones, $\{D^{(j)}(T, Y)\}_{j=1}^{l}$, by building an extended matrix or by computing their weighted average.

[3] This is a *Wholeset* method. Undoubtedly, for "large size" applications, we can select the small number of representatives from the given training data set through the clustering phase. Rather than deciding to discard or retain the training points with the *Random_C*, *PeatSeal*, or *KCentres* [13], we can do this by invoking a PRS (Prototype Reduction Scheme). For the interest of brevity, the details of the *PRS-based methods* are omitted here, but can be found in [17].

4. For any dissimilarity matrix, $D^{(j)}(T,Y)$, $(j=1,\cdots,l)$, perform classification of the input, $\boldsymbol{z}$, with *combined* classifiers designed on the newly created dissimilarity space as follows:
   (a) Compute a dissimilarity column vector, $d^{(j)}(\boldsymbol{z})$, for the input sample $\boldsymbol{z}$, with the same method as in measuring the $D^{(j)}(T,Y)$.
   (b) Classify $d^{(j)}(\boldsymbol{z})$ by invoking a group of DBCs as the *base* classifiers designed with n m-dimensional vectors in the dissimilarity space. The classification results are labeled as $class_1, class_2, \cdots, class_M$, respectively.
5. Obtain the final result from the $class_1, class_2, \cdots, class_M$ by combining the base classifiers designed in the above step, where the base classifiers are combined to form the final decision in the *fixed* or *trained* fashion.

In the above algorithm, using the $n \times n$ dissimilarity matrix, the feature-based vectors are translated into the *dissimilarity-based vectors*, where the dimensionality is determined with the number of samples n. While the dimensionality of the feature-based vectors is p, thus, the dimensionality of the dissimilarity-based vectors is $n(<< p)$. Notice also that the sample to be tested is projected onto the dissimilarity space represented by the dissimilarity matrix. From these considerations, it can be noted that the algorithm can be used as a scheme to reduce the dimensionality without encountering the SSS problem in FR.

In Step 3, on the other hand, a number of distinct dissimilarity matrices can be combined into a new one to obtain a more powerful representation in the discrimination. A simple method to do this is to average different representations. For example, two dissimilarity matrices, $D^{(1)}(T,Y)$ and $D^{(2)}(T,Y)$, can be averaged into $\frac{1}{2}(\alpha_1 D^{(1)}(T,Y) + \alpha_2 D^{(2)}(T,Y))$ after scaling with an appropriate weight, α_τ, to guarantee that they all take values in a similar range. In addition to this averaging method, the two dissimilarity matrices can be combined into : $\sum_{\tau=1}^{2} \log(1+\alpha_\tau D^{(\tau)}(T,Y))$, $\min_\tau\{\alpha_\tau D^{(\tau)}(T,Y)\}$, and $\max_\tau\{\alpha_\tau D^{(\tau)}(T,Y)\}$ [12]. Some of them will be exhaustively investigated in the present experiment.

The computational complexity of the proposed algorithm depends on the computational costs associated with the dissimilarity matrix. The time complexity of CDBC can be analyzed as follows: Step 1 requires $O(1)$ time. Step 2 requires $k \times O(n^2) = O(n^2)$ time to compute the k dissimilarity matrices. Step 3 requires $l \times O(n^2) = O(n^2)$ time to compute the l combined matrices, for example, by averaging the l matrices. Step 4 requires $O(n) + M \times O(\gamma_1) = O(\gamma_1)$ time (where M is the number of the base classifiers and γ_1 is the time for doing classification with the base classifiers.) to project the test sample onto the dissimilarity space and classify it with the *base* classifiers designed in the dissimilarity space. Step 5 requires $O(\gamma_2)$ time to classify the test sample with the *combined* classifier designed in the dissimilarity space. Here, γ_2 is the time for obtaining the final result. Thus, the total time complexity of the CDBC is $O(n^2 + \gamma_1 + \gamma_2)$. Then, the space complexity of CDBC is $O(n(n+p))$.[4]

[4] In [9], it was reported that the time complexities of LDA-extension methods such as PCA, PCA+LDA, LDA/GSVD, and RLDA, respectively, are $O(n^2p)$, $O(n^2p)$, $O((n+c)^2p)$, and $O(n^2p)$ and their space complexities are all the same as $O(np)$.

3 Experimental Results

3.1 Experimental Data

The proposed method has been tested and compared with conventional methods. This was done by performing experiments on two well-known benchmark face databases, namely, the "AT&T"[5] and "Yale"[6] databases.[7]

The face database captioned AT&T, formerly the ORL database of faces, consists of ten different images of 40 distinct subjects, for a total of 400 images. Each subject is positioned upright in front of a dark homogeneous background. The size of each image is 112×92 pixels, for a total dimensionality of 10304. The face database termed as Yale contains 165 gray scale images of 15 individuals. The size of each image is 243×320 pixels, for a total dimensionality of 77760. However, in this experiment, each facial image of 236×178 pixels was manually extracted, and then represented by a centered vector of normalized intensity values.

3.2 Experimental Method

In this paper, all experiments were performed using a "leave-one-out" strategy. To classify an image of object, that image is removed from the training set and the dissimilarity matrix is 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 dissimilarity space using the dissimilarity matrix, and recognition was performed based on the proposed algorithm in Section 2.3. We repeated this n times for every sample and obtained a final result by averaging them.

To construct the dissimilarity matrix, all samples were selected as representatives and the dissimilarities were measured with the Euclidean distance and the regional distance. Here the two distance measures are named as "ED" and "RD", respectively.[8] The distance measure called RD is defined as the average of the minimum difference between the gray value of a pixel and the gray value of each pixel in the 5×5 neighborhood of the corresponding pixel. In this case, the regional distance compensates for a displacement of up to three pixels of the images. For the interest of brevity, the details of the distance measure are omitted here, but can be found in the literature including [19].

However, the faces for some subjects vary with pose, illumination, facial expression, and whether or not they are wearing glasses. Thus, the dissimilarity

[5] http://www.cl.cam.ac.uk/Research/DTG/attarchive/facedatabase.html

[6] http://www1.cs.columbia.edu/ belhumeur/pub/images/yalefaces

[7] A thorough evaluation on AT&T and Yale databases is presented here. It would be interesting to see results on more challenging datasets, such as FERET and CMU-PIE. The results on these datasets will appear in the next paper.

[8] Here, we experimented with two simple measures, namely, ED and RD. However, it should be mentioned that we can have numerous solutions, depending on dissimilarity measures, such as the Hamming distance, the modified Hausdorff distances, the blurred Euclidean distance, etc. From this perspective, the question "what is the best measure?" is an interesting issue for further study.

matrix simply obtained by measuring the input images can not work as a representative. To overcome this problem as well as the SSS problem, a combined dissimilarity representation and two classifier fusion strategies are employed in the experiment. To investigate this combination rule, first of all, two dissimilarity representations, namely, ED and RD, are averaged into a new representation (which is named as "AD" here) after normalization. As mentioned in the previous section, *three* base classifiers are designed in this newly defined dissimilarity space, and then all of their results are combined in *fixed* or *trained* fashion.

Since the diversity between the base classifiers is essential for constructing a robust ensemble, different classifiers, such as Nearest Mean Classifiers, Normal Density based Classifiers, and Nonlinear Classifiers, are considered as the base classifiers. These three kinds of base classifiers are implemented with PRTools,[9] and will be denoted as *nmc*, *ldc*, and *knnc*, respectively, in a subsequent section. The outputs of the base classifiers are combined with fixed combiners, such as *Product*, *Median*, and *Majority vote* rules, and two trained classifiers. All *five* combiners are also implemented with PRTools, and named as *prodc*, *medianc*, *votec*, *meanc*, and *fisherc*, respectively. To simplify the classification task for the paper, only three base classifiers, three fixed and two trained combiners are experimented. However, other classifiers, including neural network and SVM based classifiers, and combining rules can also be considered.

3.3 Experimental Results

The run-time characteristics of the proposed algorithm for the two benchmark databases, AT&T and Yale, is reported below and shown in Table 1. The performance of the dissimilarity-based classifiers (DBC and CDBC) is investigated first. Following this, a comparison is made between the conventional LDA-extension methods and the proposed CDBC scheme.

First of all, to examine the rationality of employing a fusion technique in the CDBC, the simple Dissimilarity-Based Classifier (DBC) was experimented. While CDBC involves all of the five steps given in Section 2.3, DBC consists of only the steps 1, 2, and 4 with $k = 1$ and $l = 1$. The classification accuracy rates of DBC was evaluated for the AT&T and Yale databases. In this experiment, the same dissimilarity matrix was constructed for both DBC and CDBC.

Table 1 shows the classification accuracy rates (%) of DBCs and CDBCs for the two databases. Here, the abbreviations ED, RD, and AD, which are the Eucledian distance, the regional distance, and the averaged distance, indicate the dissimilarity measures employed in this experiment. Additionally, in the base classifiers column, an Uncorrelated Normal based Quadratic Classifier (named as *udc*) was used for the RD representation instead of the Normal Density based Classifier (*ldc*). Also, *knnc* stands for the k-Nearest Neighbor Classifier ($k = 1$).

From Table 1, it is observed that the classification accuracies for the benchmark databases can be improved by employing the philosophy of CDBC. This is clearly shown in the classification accuracy rates of the classifiers designed for

[9] PRTools is a Matlab Toolbox for Pattern Recognition. PRTools can be downloaded from the PRTools website, http://www.prtools.org/

Table 1. A comparison of classification accuracy rates (%) of the base Dissimilarity-Based Classifiers (DBC) and the Combined Dissimilarity-Based Classifiers (CDBC) designed with the fixed and the trained combiners. Here, the classifiers of *nmc*, *ldc (udc*)*, and *knnc* are designed and evaluated as DBCs. Then, the combiners of *prodc*, *medianc*, and *votec* are employed as the fixed combining schemes of the DBCs. Finally, the classifiers of *meanc* and *fisherc* are employed as the trained combiners respectively.

Data Sets	Distance Measures	Base Classifiers			Fixed Combiners			Trained Combiners	
		nmc	*ldc(udc*)*	*knnc*	*prodc*	*medianc*	*votec*	*meanc*	*fisherc*
AT&T	ED	81.25	98.75	96.50	98.75	98.25	98.00	98.75	99.00
	RD*	71.25	88.00	95.00	88.00	89.25	89.00	88.00	89.00
	AD	76.25	99.25	95.75	99.25	98.50	98.00	99.25	99.25
Yale	ED	80.61	93.33	79.39	93.33	86.06	86.06	93.33	93.33
	RD*	78.18	72.12	79.39	72.12	76.36	76.97	72.12	78.18
	AD	79.39	96.36	78.79	95.76	82.42	86.67	96.36	96.36

the AT&T database measured with ED. Specifically, the classification accuracies of the base classifiers, namely, *nmc*, *ldc*, and *knnc*, are 81.25, 98.75, and 96.50 (%), while those of the fixed combiners, such as *prodc*, *medianc*, and *votec*, are 98.75, 98.25, and 98.00 (%), respectively. Additionally, the trained combiners of *meanc* and *fisherc* have the classification accuracies of 98.75 and 99.00 (%), respectively. From this consideration, it is evident that the rationale of the paper for employing a fusion technique works well. Furthermore, the result of the comparison is completely in accord with the well-known fact that the combination of different classifiers for the same feature set only slightly improves the best individual results. Besides this, the results also prove that the best overall result is obtained by a trained combiner. This is the case of the *fisherc* here. For the Yale database, the same characteristics can be observed.

Secondly, as the main results, it should be noted that it is possible to improve the classification performance by appropriately combining the dissimilarity representations. For instance, the classification accuracy rates of the three base classifiers designed with AD for AT&T database are (76.25, 99.25, 95.75) (%), respectively, and those of the three fixed and the two trained combiners applied to the outputs of the base classifiers are (99.25, 98.50, 98.00) and (99.25, 99.25) (%), respectively. The above comparison shows that the accuracy rates of the combiners are generally higher than those of the base classifiers. From these considerations, the reader should observe that the newly created dissimilarity representation of AD improves the performance of the classification accuracy more effectively than the ED or RD measure. Therefore, it can be concluded that the highest accuracy rates are achieved when the combined representation, namely, AD, is classified with the trained combiners. However, it should be also pointed out that the classification efficiencies were not improved in both combiners for RD. For the Yale database, the same characteristics can be observed.

Finally, CDBC can be compared with LDA-extensions for solving the SSS problem in FR. Consider experimental results on the LDA-extensions, such as the PCA [3], the PCA+LDA [3], the direct LDA [8], the DCV [10], and the

LDA/GSVD [5], which have been recently reported in [18]. In that experiment, to reduce the computational complexity, each facial image from the two databases, AT&T and Yale, was down-sampled into 56×46 and 61×80, respectively. Additionally, the "leave-one-out" strategy was also used to experiment with these methods. In [18], the classification accuracies of the PCA, PCA+LDA, direct LDA, DCV, and LDA/GSVD methods for AT&T and Yale databases are (93.25, 95.50, 98.50, 97.25, 93.50) and (72.73, 74.55, 92.12, 70.91, 98.79) (%), respectively. A comparison of these figures and Table 1 shows that the classification accuracy of CDBCs is marginally higher than that of the conventional methods. From this consideration, the rationale of the dissimilarity-based scheme for employing a fusion technique is proven to be valid.

In review, it is not easy to say that one specific method is superior to others for solving the SSS problem in FR. However, as a matter of comparison, it is clear that the combined dissimilarity-based method is better than the conventional schemes with regard to the classification accuracy rates.

4 Conclusions

In this paper a method that seeks to address the SSS problem of image recognition by combining the dissimilarity-based classifiers was considered. Rather than use Fisher's criterion to reduce the dimensionality, a completely different approach was employed, in which an object was represented based on the dissimilarity measures among training samples instead of the feature vector itself. Apart from utilizing the dissimilarity representation [18], to increase the classification accuracy, in this paper, a new method of employing a fusion technique in representing features as well as in designing classifiers was proposed.

The proposed method has been tested on two well-known face databases and compared with LDA-extension approaches. The experimental results demonstrate that the proposed scheme works well and its classification accuracy is better than that of the conventional ones. The results especially demonstrate that the highest accuracy rates are achieved when the combined representation is classified with the trained combiners. Although an investigation was made that focused on the possibility that the combined dissimilarity-based classifiers could be used to solve the SSS problem, many problems remain. One of them is an improvement of the classification performance by utilizing an appropriate dissimilarity measure (i.e., a modified Hausdorff distance) and by developing a suitable feature combination in the dissimilarity space. The research concerning this is a future aim of the authors.

References

1. W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips, "Face recognition: a literature survey", *ACM Compt. Surveys*, vol. 35, no. 4, pp. 399 - 458, 2003.
2. J. Ruiz-del-Solar and P. Navarrete, "Eigenspace-based face recognition: a comparative study of different approaches", *IEEE Trans. Systems, Man, and Cybernetics - Part C*, vol. 35, no. 3, pp. 315 - 325, Aug. 2005.

3. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 19, no. 7, pp. 711 - 720, July 1997.
4. L. -F. Chen, H. -Y. M. Liao, M. -T. Ko, J. -C. Lin, and G. -J. Yu, "A new LDA-based face reognition system which can solve the small sample size problem", *Pattern Recognition*, vol. 33, pp. 1713 - 1726, 2000.
5. P. Howland, J. Wang, and H. Park, "Solving the small sample size problem in face reognition using generalized discriminant analysis", *Pattern Recognition*, vol. 39, pp. 277 - 287, 2006.
6. S. Raudys and R. P. W. Duin, "On expected classification error of the Fisher linear classifier with pseudoinverse covariance matrix", *Pattern Recognition Letters*, vol. 19, pp. 385 - 392, 1998.
7. D. Q. Dai and P. C. Yuen, "Regularized discriminant analysis and its application to face recognition", *Pattern Recognition*, vol. 36, pp. 845 - 847, 2003.
8. H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data - with application to face recognition", *Pattern Recognition*, vol. 34, pp. 2067 - 2070, 2001.
9. J. Ye and Q. Li, "A two-stage linear discriminant analysis via QR-decomposition", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 27, no. 6, pp. 929 - 941, Jun. 2005.
10. H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 27, no. 1, pp. 4 - 13, Jan. 2005.
11. E. Pekalska and R. P. W. Duin, "Dissimilarity representations allow for buiilding good classifiers", *Pattern Recognition Letters*, vol. 23, pp. 943 - 956, 2002.
12. E. Pekalska, *Dissimilarity representations in pattern recognition. Concepts, theory and applications*, Ph.D. thesis, Delft University of Technology, The Netherlands, 2005.
13. E. Pekalska, R. P. W. Duin, and P. Paclik, "Prototype selection for dissimilarity-based classifiers", *Pattern Recognition*, vol. 39, pp. 189 - 208, 2006.
14. L. I. Kuncheva, "A theoretical study on six classifier fusion strategies", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 24, no. 2, pp. 281 - 286, Feb. 2002.
15. J. Kittler, M. Hatef, R. P. W. Duin and J. Matas, "On combining classifiers", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 20, no. 3, pp. 226 - 239, Mar. 1998.
16. L. I. Kuncheva, J. C. Bezdek and R. P. W. Duin "Decision templates for multiple classifier fusion: an experimental comparison", *Pattern Recognition*, vol. 34, pp. 299 - 414, Feb. 2001.
17. S. -W. Kim and B. J. Oommen, "On optimizing dissimilarity-based classification using prototype reduction schemes", *The 2006 International Conference on Image Analysis and Recognition (ICIAR2006)*, LNCS 4141, pp. 15-28, 2006.
18. S. -W. Kim, "On solving the small sample size problem using a dissimilarity representation for face recognition", *Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS2006)*, LNCS 4179, pp. 1174-1185, 2006.
19. Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 19, no. 7, pp. 721 - 732, Jul. 1997.