

# H-005 On Using DTW and Multiple Fusion Strategy for Optimizing Dissimilarity-Based Classification

Sang-Woon Kim<sup>†</sup> and Seunghwan Kim<sup>‡</sup>

<sup>†</sup>Dept. of Computer Engineering, Myongji University, Korea

<sup>‡</sup>School of Electrical Engineering, University of Waterloo, Canada

## 1 Introduction

One of the most recent and novel developments in pattern classification is the concept of dissimilarity-based classifiers (DBC) proposed by Duin and his co-authors [1]. DBCs are a way of defining classifiers between the classes, which are not based on the feature measurements of the individual patterns, but rather on a suitable *dissimilarity measure* between them [1]. In this strategy, therefore, we need to measure the inter-pattern dissimilarities for all the training samples to ensure there is no zero distance between objects of different classes. In image classification tasks, such as face recognition, one of the most intractable problems is the distortion and lack of information caused by the differences in face directions and sizes. To overcome these problems, we employ a way of measuring the dissimilarity distance between two images of an object using a dynamic programming technique, such as dynamic time warping [2]. 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 [3]. Thus, to increase the classification accuracy of DBCs further, we also use a method of simultaneously employing multiple fusion strategies in representing features as well as in designing classifiers [4].

## 2 Dissimilarity-Based Classifications

A dissimilarity representation of a set of samples,  $T = \{\mathbf{x}_i\}_{i=1}^n \in \mathcal{R}^d$ , is based on pairwise comparisons and is expressed, for example, as an  $n \times m$  dissimilarity matrix  $D_{T,Y}[\cdot, \cdot]$ , where  $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_m\}$ , a prototype set, is extracted from  $T$ , and the subscripts of  $D$  represent the set of elements, on which the dissimilarities are evaluated. Thus, each entry  $D_{T,Y}[i, j]$  corresponds to the dissimilarity between the pairs of objects  $\langle \mathbf{x}_i, \mathbf{y}_j \rangle$ , where  $\mathbf{x}_i \in T$  and  $\mathbf{y}_j \in Y$ . Consequently, an object  $\mathbf{x}_i$  is represented as a column vector as follows:

$$[d(\mathbf{x}_i, \mathbf{y}_1), d(\mathbf{x}_i, \mathbf{y}_2), \dots, d(\mathbf{x}_i, \mathbf{y}_m)]^T, 1 \leq i \leq n. \quad (1)$$

Here, the dissimilarity matrix  $D_{T,Y}[\cdot, \cdot]$  is defined as a *dissimilarity space*, on which the  $d$ -dimensional object,  $\mathbf{x}$ , given in the feature space, is represented as an  $m$ -dimensional vector  $\delta_Y(\mathbf{x})$ .

To compute the dissimilarity matrix, we first select the representatives using a prototype selection method, such as *Random*, *RandomC*, *KCentres*, *ModeSeek*, and so on, or using all training samples as the representative.

Then, we measure the dissimilarities between them using the measuring systems, such as Euclidean distance, Hamming distance, the regional distance, and the spatially weighted gray-level Hausdorff distance measures. The details of the DBCs are omitted here in the interest of compactness, but can be found in the existing literature, including [1].

## 3 DTW and Fusion Strategies

With regard to measuring the dissimilarity of the sample points, we prefer not to directly measure the dissimilarity from the object points; rather, we utilize a way of using the DTW (dynamic time warping) technique to *adjust* or *scale* the object samples. This measure of dissimilarity effectively serves as a new “feature” component in the dissimilarity space.

Consider the two sequences of  $s = (x_1, \dots, x_n) \in T$  and  $t = (x_1, \dots, x_m) \in T$ , where  $x_i$  is an element that corresponds to a column vector of an image sample at index  $i$  in an  $s$  or  $t$  sample. An alignment from  $s$  to  $t$  can be represented by a warping  $w = \{w(1), w(2), \dots, w(n)\}$ , where  $j = w(i), j \in [1, m], i \in [1, n]$  means that the  $i$ -th element in  $s$  is aligned to the  $j$ -th element in  $t$ .

To find the best warping path  $w$  that minimizes the distance  $D_w(s, t)$ , we use the correlation coefficient between two (sub)vectors  $x_i$  and  $x_j$ ,  $\rho(x_i, x_j)$ , which is defined as:  $\rho(x_i, x_j) = \frac{\text{cov}(x_i, x_j)}{\sigma_{x_i} \sigma_{x_j}} = E((x_i - \mu_{x_i})(x_j - \mu_{x_j})') / \sigma_{x_i} \sigma_{x_j}$ , where  $\mu_{x_i}$  and  $\sigma_{x_i}$  are expected values and standard deviations, respectively. Also  $E$  is the operator for expected values.

To combine dissimilarity matrices obtained with different measuring systems to a new representation matrix, we make use of representation combining strategies, such as *Average*, *Product*, *Min*, and *Max* rules. For example, in the *Average* rule, two dissimilarity matrices,  $D^{(1)}(T, Y)$  and  $D^{(2)}(T, Y)$ , can be averaged to  $(\alpha_1 D^{(1)}(T, Y) + \alpha_2 D^{(2)}(T, Y))$  after scaling with an appropriate weight,  $\alpha_i$ , to guarantee that they all take values in a similar range. The details of the other methods are omitted here, but can be found in [1].

## 4 Schema for the Proposed Solution

To solve the classification problem, we first combine dissimilarity matrices constructed with various measuring systems including the dynamic time warping for the entire training samples, and then again combine all of the results of DBCs designed on the combined dissimilarity space to reduce the classification error rates. The

proposed algorithm for the combined DBCs is summarized in the following:

1. Select the entire training samples  $T$  as the representative set  $Y$ .

2. Using Eq. (1), compute dissimilarity matrices,  $D^{(1)}(T, Y)$ ,  $D^{(2)}(T, Y)$ ,  $\dots$ ,  $D^{(k)}(T, Y)$ , by using the  $k$  different dissimilarity measures for all  $x \in T$  and  $y \in Y$ .

3. For any  $D^{(j)}(T, Y)$ , ( $j = 1, \dots, l$ ), perform classification of the input,  $z$ , with *combined* classifiers designed on the combined dissimilarity space as follows:

3.1 Compute a dissimilarity column vector,  $\delta^{(j)}(z)$ , for the input sample  $z$ , with the same method as in measuring the  $D^{(j)}(T, Y)$ .

3.2 Classify  $\delta^{(j)}(z)$  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  $class_1, class_2, \dots$ , respectively.

4. Obtain the final result from the  $class_1, class_2, \dots$ , 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.

The rationale of this strategy is presented in a later section together with the experimental results.

## 5 Experimental Results

The proposed method has been tested and compared with conventional methods. This was done by performing experiments on three well-known benchmark databases, namely AT&T (shortly, A) <sup>1</sup>, Yale (shortly, Y) <sup>2</sup>, and RoadSign (shortly, R) [5].

We first combined two dissimilarity matrices measured with Euclidean distance (ED) and dynamic time warping (DTW) techniques to a new representation matrix (FUD). Then we combined again all of the results of the base classifiers, which had been trained in the new representation matrix, in *fixed* or *trained* fashion.

Three base classifiers denoted as *nmc*, *ldc*, and *knnc* were implemented with PRTools <sup>3</sup>. Then, two combiners (a fixed and a trainable) were also implemented with PRTools and named as *prodc* and *meanc*, respectively. Table 1 shows a comparison between the classification accuracy rates (%) of combined DBCs for the three databases.

From Table 1, it is clear that there is an improvement in the achieved classification accuracies. An example of this is the classification accuracies of the *meanc* for AT&T (A) database. For the *three* classifiers trained in the three dissimilarity spaces, namely ED, DTW, and FUD, the classification accuracy rates are 95.10(%), 89.40(%), and 97.00(%), respectively. The same characteristic could also be observed from the other databases, such as Yale and RoadSign. From this consideration, the reader can observe that the classification performances of DBCs trained in FUD are usually better than those of DBCs built in ED and DTW spaces, which leads to

Table 1: A comparison of the classification accuracy rates (%) of combined DBCs trained with the Euclidean (ED), dynamic time warping (DTW), and fusion (FUD) based methods for the three databases.

data sets	dis-space	base classifiers			combiners	
		<i>nmc</i>	<i>ldc</i>	<i>knnc</i>	<i>prodc</i>	<i>meanc</i>
A	ED	73.900	95.100	89.000	95.100	95.100
	DTW	81.700	89.400	92.400	89.400	89.400
	FUD	77.200	97.000	90.300	97.000	97.000
Y	ED	63.333	80.667	69.667	80.667	80.667
	DTW	65.444	80.000	70.667	80.222	80.111
	FUD	64.111	81.444	70.222	81.333	81.444
R	ED	99.948	99.991	99.982	99.991	99.991
	DTW	99.948	99.993	99.985	99.993	99.992
	FUD	99.948	99.994	99.984	99.994	99.994

the conclusion that combining dissimilarity matrices is helpful. It is also interesting to point out that the scaling factors,  $\alpha_1:\alpha_2$ , used for the AT&T, Yale, and RoadSign databases are 0.75 : 0.25, 0.35 : 0.65, and 0.7 : 0.3, respectively. Thus, the problem of automatically selecting an optimal scaling factor  $\alpha_i$  for a given application remains unresolved.

## 6 Conclusion

In this paper we studied a method of using the dynamic programming and multiple fusion strategies for optimizing dissimilarity-based classification (DBC). The proposed scheme was experimented and compared with the conventional methods for the well-known benchmark image databases. Our experimental results demonstrated the possibility that the proposed method could be used efficiently for optimizing DBCs. The research concerning the selection of an optimal scaling factor is a future aim of the authors.

**Acknowledgments** This work was supported by the Korea Research Foundation Grant funded by the Korean Government(KRF-2008-013-D00115). This paper was partially presented at the IEEK Summer Conference 2009, in Jeju, Korea, July 2009.

## References

- [1] E. Pekalska, R. P. W. Duin, *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*, World Scientific Publishing, Singapore (2005).
- [2] S. -W. Kim, J. Gao, A dynamic programming technique for optimizing dissimilarity-based classifiers, *Proc. of SSSPR'08*, LNCS-5342 664–673 (2008).
- [3] R. I. Kuncheva, *Combining Pattern Classifiers - Methods and Algorithms*, John Wiley & Sons, New Jersey (2004).
- [4] S. -W. Kim, R. P. W. Duin, On optimizing dissimilarity-based classifier using multi-level fusion strategies, *Journal of Institute of Electronics Engineers of Korea*, 45-CI(5) 15–24 (2008) (*In Korean*).
- [5] P. Paclik, J. Novovicova, P. Somol, P. Pudil, Road sign classification using Laplace kernel classifier, *Patt. Recogn. Lett.*, 21(13-14) 1165–1173 (2000).

<sup>1</sup>[http://www.cl.cam.ac.uk/Research/DTG/attarchive/face\\_database.html](http://www.cl.cam.ac.uk/Research/DTG/attarchive/face_database.html)

<sup>2</sup><http://www1.cs.columbia.edu/belhumeur/pub/images/yale-faces>

<sup>3</sup>PRTools is a Matlab toolbox for pattern recognition.