Collaborative Sparse Representation in Dissimilarity Space for Classification of Visual Information

Ilias Theodorakopoulos, George Economou and Spiros Fotopoulos

Electronics Laboratory, Department of Physics, University of Patras, Patras, Greece
iltheodorako@upatras.gr,
{spiros,economou}@physics.upatras.gr

Abstract. In this work we perform a thorough evaluation of the most popular CR-based classification scheme, the SRC, on the task of classification in dissimilarity space. We examine the performance utilizing a large set of public domain dissimilarity datasets mainly derived from classification problems relevant to visual information. We show that CR-based methods can exhibit remarkable performance in challenging situations characterized by extreme non-metric and non-Euclidean behavior, as well as limited number of available training samples per class. Furthermore, we investigate the structural qualities of a dataset necessitating the use of such classifiers. We demonstrate that CR-based methods have a clear advantage on dissimilarity data stemming from extended objects, manifold structures or a combination of these qualities. We also show that the induced sparsity during CR, is of great significance to the classification performance, especially in cases with small representative sets in the training data and large number of classes.

1 Introduction

Dissimilarity representations are commonly used in several contemporary applications, where the structural approach to intelligent data analysis is followed as a natural and powerful alternative to the traditional vectorial representations. During the previous decades many techniques aiming to quantify the dissimilarity between objects were developed. The problem of computing a dissimilarity value can be viewed from different perspectives: using dynamic programming in order to estimate the non-linear warping between two trajectories in the feature space [1], transforming the problem into the graph domain and comparing the corresponding graphs [3], considering the given sets of vectors as multivariate distributions and applying statistical tests [11] or computing appropriate distances [8,16] etc. Regardless of the method chosen, the resultant dissimilarity values are the quantification of pairwise object comparisons in a global manner. Additionally, such a representation provides a context where information fusion can be expressed very easily (i.e. averaging dissimilarity values derived from different descriptors).

In this context, a large number of applications aim to classify objects using pairwise dissimilarities to a set of labeled objects. In order to utilize modern classifiers, the dissimilarity data typically has to be transformed into a vector representation

G. Bebis et al. (Eds.): ISVC 2013, Part I, LNCS 8033, pp. 496–506, 2013.

through embedding into a vector space of fixed dimensions. The problem is that usually the dissimilarity measures and the resultant data do not satisfy the mathematical requirements of a metric function, that is the underlying Gram matrix to be positive semi-definite. Consequently, the dissimilarity data have to be further processed in order to suppress the non-Euclideanity and/or non-metricity. Some popular methods for transforming non-Euclidean dissimilarity into vector representations include the positive definite subspace embedding i.e. classical multidimensional scaling [27], based on the assumption that metric violations are noise artifacts lacking of useful information, the pseudo-Euclidean embedding [12] , generalizations such as MDS [4] and kernel PCA [24], and the manifold embedding i.e. [28], assuming that Euclidean violations are an outcome of the intrinsic manifold structure of the data.

Recent studies [18, 5] though, have shown that significant discriminative information can be expressed through the negative eigenspace. This particularly stands in applications related to the perception of visual information, where the raw data are usually characterized by very high dimensionality compared to that of the underlying structure of the classes. Additionally, the human perception of dissimilarity between objects is rarely Euclidean and commonly non-metric, resulting a considerable amount of information to be encoded in the negative eigenspace. Therefore, although a highly non-Euclidean/ non-metric measure is able to describe a given problem quite well, data cannot be embedded distortionless into a real Euclidean space.

An alternative approach to the embedding into a vector-space is the representation in the dissimilarity space [21], where each sample is directly represented by a vector of dissimilarities with a set of representative samples. In this case, the information lying in the negative eigenspace is preserved since no modification is being performed to the data. The properties of such spaces have not been extensively studied yet, although there is a solid justification for the construction of classifiers in dissimilarity spaces [20]. Furthermore, there is strong evidence regarding to the benefits of using classifiers in this context, as recently reported by R. Duin et al. [5], achieving state-of-the-art classification performance on publicly available dissimilarity data, using linear SVM in the dissimilarity space. Furthermore, an extensive evaluation [6] of several linear classifiers operating in the dissimilarity space, including SVMs, Fisher discriminant, Linear Logistic regression etc., revealed that they perform similar or better compared to linear and non-linear feature-based classifiers in a wide range of datasets.

Motivated by these findings, several works have been published recently, investigating different aspects of dissimilarity-based pattern recognition. Hammer et al. [15] proposed a scheme for prototype-based classification of possibly non-Euclidean dissimilarity data. Schleif et al. [23] proposed a prototype-based conformal classifier which enables the calculation of a confidence measure for the produced classification. Both of these schemes are based on the relational prototype based learning, where is assumed that the prototypes are linear combination of the underlying data points. Elhamifar et al. [7] proposed a scheme for the discovery of appropriate exemplars, in order to efficiently represent the data using non-Euclidean/non-metric dissimilarities. They formulated the problem as a row-sparsity trace minimization problem, solved via convex programming. Calana et. al. [22] proposed a supervised criterion for the

selection of feature lines, a concept associated with generalized dissimilarity representations which has proven to be very efficient for small representation sets. A similar approach has also been proposed [25], presented as a two-stage, sparse-based classification scheme for the human action recognition task, where at the first stage a test sequence is represented as a mixture of the training actions and at the second stage this mixture is used in order to classify the sequence.

In this work we aim to thoroughly investigate the capabilities of collaborative sparse representation regarding the classification in the dissimilarity space, and especially the discriminative efficiency in challenging visual classification tasks. This classification scheme is based on the popular SRC [29] classifier, and has recently been successfully utilized [26] for the task of human action recognition using pose data from low-cost devices, characterized by significant inaccuracies and noise. Furthermore, we aim to define the structural characteristics of a dataset that constitute the utilization of the above classification scheme truly beneficial. The rest of this work is organized as follows: The basic formulation of representation in the dissimilarity space is given in section 2. The principles and basic properties of collaborative representation-based classification are detailed in Section 3. The thorough experimental procedure, and the obtained results are given in Section 4. Conclusions are drawn in Section 5.

2 2Representation in Dissimilarity Space

Let $\mathbf{S}=\{o_1,o_2,...,o_n\}$ be a training set of objects o_i, represented by an arbitrary type of data. Given an appropriate dissimilarity function d, a mapping $\mathbf{X}(\bullet,P):\mathbf{S}\to\mathbb{R}^k$ can be defined, where $P=\{p_1,p_2,...,p_k\}$ is, in the general case, a set of k objects, namely prototypes, and can be a subset of $\mathbf{X}$. In the resulting space, called dissimilarity space, each dimension $\mathbf{X}(\bullet,p_i)$ describes the dissimilarity to the i^{th} prototype. In the current work we assume that $P:=\mathbf{S}$, so as every object to be represented by an n-dimensional vector of dissimilarities to all training objects :

$$\mathbf{y}=\mathbf{X}(o,\mathbf{S})=\left[d(o,o_1),d(o,o_2),...,d(o,o_n)\right]^{\mathrm{T}} \tag{1}$$

Thereafter, the representative vectors for the objects of $\mathbf{S}$ are simply the columns of the corresponding dissimilarity matrix $\mathbf{X}$.

3 Classification Based on Collaborative Representation

Collaborative representation (CR) constitutes the representation of a data sample $\mathbf{y}\in\mathbb{R}^m$ as a linear combination $\mathbf{y}\approx\mathbf{Xa}$ of n training samples from K classes forming the dictionary $\mathbf{X}\in\mathbb{R}^{m\times n}$. Thus, $\mathbf{X}$ is of the form $\mathbf{X}=[\mathbf{X}_1,\mathbf{X}_2,...\mathbf{X}_K]$, where $\mathbf{X}_i$ is the matrix holding the training samples from the i^{th} as column vectors. The coefficient

vector $\mathbf{a} \in \mathbb{R}^n$ of the representation is computed by solving an optimization problem of the form

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_p \tag{2}$$

where λ is a regularization factor. The classification of the test sample $\mathbf{y}$ can now be performed by seeking for the partial linear combination $\mathbf{X}_i\hat{\mathbf{a}}_i$, where $\hat{\mathbf{a}}_i$ is the vector holding the coefficients associated with the training samples of class *i*, that best represents the test sample $\mathbf{y}$ in terms of minimum reconstruction error. In a more formal way

$$Identity(\mathbf{y}) = \arg\min_i \|\mathbf{y} - \mathbf{X}_i\hat{\mathbf{a}}_i\|_2^2 \tag{3}$$

In order to gain some insights regarding the classification mechanism based on collaborative representation, we can discard the regularization factor from the optimization problem (1), for simplicity. Thus, the representation becomes the least-square problem $\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2^2$. The associated representation $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{a}}$ is the perpendicular projection of $\mathbf{y}$ onto the space spanned by $\mathbf{X}$. The reconstruction error by each class can be written

$$e_i = \|\mathbf{y} - \mathbf{X}_i\hat{\mathbf{a}}\|_2^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 + \|\hat{\mathbf{y}} - \mathbf{X}_i\hat{\mathbf{a}}_i\|_2^2 \tag{4}$$

The useful term for discrimination is the $e_i^* = \|\mathbf{y} - \mathbf{X}_i\hat{\mathbf{a}}\|_2^2$, since the amount $\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2$ is constant for all classes. It can be easily shown [28] that the representation error can be represented as

$$e_i^* = \frac{\sin^2(\hat{\mathbf{y}}, \chi_i)\|\hat{\mathbf{y}}\|_2^2}{\sin^2(\chi_i, \bar{\chi}_i)} \tag{5}$$

where $\chi_i = \mathbf{X}_i\hat{\mathbf{a}}_i$ is the projection of $\mathbf{y}$ onto the subspace spanned by the samples of class i , and $\bar{\chi}_i = \sum_{i \neq j} \mathbf{X}_j\hat{\mathbf{a}}_j$ is the projection onto the subspace spanned by the samples of all the other classes. From eq. (4) it is apparent that the minimization of the representation error as expressed by eq. (2), is equivalent to the quest for small angle between the overall representation $\hat{\mathbf{y}}$ and the class-depended representation χ_i and simultaneously large angle between χ_i and $\bar{\chi}_i$. Thus, when applying the classification rule (2) on collaborative representations, we seek for the class that best represents a test sample in a voracious manner. This form of dual objective gives robustness to the CR-based classification methods, especially when the classes are hardly distinguishable.

Maybe the most successful CR-based classification scheme is the Sparse Representation-based Classification scheme (SRC), proposed by Wright et al. in [29], where l_1 –norm regularization (p=1) has been imposed to the optimization problem (1), in

order to induce sparsity in the representation coefficients $\hat{\mathbf{a}}$. The role of sparsity is essential in order to make the solution of (1) stable, especially in applications with large number of classes and/or small number of training samples per class. In order to fully benefit from the induced sparsity, the dictionary $\mathbf{X}$ has to be over-complete. Therefore, PCA or random projections are usually applied to the data in order to reduce dimensionality and render the training set to an overcomplete dictionary. The problem (1) can be solved efficiently using linear programming techniques. The method has been applied in several challenging classification problems such as face recognition, achieving state-of-the-art performance. A very effective validation criterion has also been proposed, based on the class-specific sparsity of the coefficients $\hat{\mathbf{a}}_i$. In this paper we evaluate SRC classifier on several dissimilarity datasets, in order to examine the effectiveness of the CR-based classification, on the task of classifying data into dissimilarity space.

4 Experiments

In this work, our goal is to systematically assess the performance of CR-based methods on the classification of dissimilarity data. To this purpose, we evaluate SRC using a set of 11 public domain dissimilarity datasets, shown in Table 1. The utilized datasets are derived mostly from problems concerning classification of visual information both static (such as shape gradient etc.) and dynamic (human motion etc.). The first 6 datasets of Table 1 are available at the D3.3 deliverable of the EU SIMBAD project, and are considered as a benchmark in classification of dissimilarity data. Specifically, WoodyPlants is a subset of the shape dissimilarities between leaves of woody plants [17], including only classes with more than 50 objects. Catcortex is based on the connection strength between 65 cortical areas of a cat, [13]. GaussM1 and GaussM02 are based on two 20-dimensional normally distributed sets of objects, for which dissimilarities are computed using the Minkowsky distance of order 1 and 0.2 respectively. The three Coil dataset is based on the same sets of SIFT points in COIL images compared using graph distance. The Delft dataset consists of the dissimilarities computed from a set of gestures in a sign-language study [19]. They are measured by two video cameras observing the positions the two hands in 75 repetitions of creating 20 different signs. The dissimilarities result from a dynamic time warping procedure.

The remaining five consists of a group of challenging datasets, namely UPCV-Dissim, emerged from popular computer vision problems. Specifically, UPCV-Gait datasets consists of the dissimilarity matrix of pose sequences derived by capturing the gait of 22 individuals, five times each. The dissimilarities were computed using the multivariate extension of Wald–Wolfowitz statistical test. For details regarding the dataset and the dissimilarities calculation see [25]. The UPCV-Act-m and UPCV-ActD datasets correspond to the dissimilarities between sequences of poses, captured from 10 individuals performing 10 actions, twice each. The dissimilarities were computed using MNPD and DTW algorithms respectively. For details the reader can refer to [26]. The MPEG7-Shape dataset corresponds to the dissimilarities between 69

classes of shapes from the MPEG7 shape dataset. The dissimilarities were computed using time delay embedding of the Centroid Contour Descriptor and the MNPD algorithm. For details the reader can refer to [9]. Finally, the Leafs dataset consists of the dissimilarities computed on a set of leaf images, subset of a larger Herbarium database, used in [10]. It is derived from images from 37 leaf species with 25 leaves from each species. The dissimilarities resulted according to the authors using multivariate extension of Wald–Wolfowitz statistical test via mapping of the angle sequence (AS) descriptor of leaf contours into the phase space.

In the first section of Table 1 there are some general properties of the datasets: number of objects (size), number of classes, the fraction of triangle violations in the corresponding dissimilarity matrix (non-metric) and the Negative Eigen-Fraction (NEF). As can be seen, the utilized datasets span across a wide range of properties, regarding non-metric and non-Euclidean behavior as also samples-per-class ratio.

Table 1. Datasets characteristics and classification errors for SRC and SVM using leave-one-out cross validation

	Dataset Characteristics				Classification Error	
Dataset	Size	Number of Classes	Non-Metric	NEF	SRC	SVM
WoodyPlants50	791	14	5E-04	0.23	**0.073**	0.075
CatCortex	65	4	2E-03	0.21	**0.015**	0.046
GaussM1	500	2	0	0.26	**0.182**	0.202
GaussM02	500	2	5E-04	0.39	**0.18**	0.204
CoilDelftSame	288	4	0	0.03	**0.406**	0.413
Delft	1500	20	9E-06	0.31	**0.025**	0.027
UPCV-Gait	110	22	6E-03	0.1	**0.046**	0.391
UPCV-Act-m	200	10	1E-03	0.13	**0.045**	0.05
UPCV-ActD	200	10	1E-02	0.21	**0.09**	0.1
MPEG7-Shape	1380	69	3E-03	0.21	**0.065**	0.107
Leafs	925	37	3E-02	0.31	**0.088**	0.102

We compare the performance of the CR-based classification to the results obtained by the linear SVM operating in the dissimilarity space, reported by Duin et al. in [5]. In order to constitute the results directly comparable, we use the same preprocessing and experimental protocol, followed in [5]. Thus, for every dataset the dissimilarity matrix is made symmetric by averaging with its transpose and normalized by the average off-diagonal dissimilarity. Error estimates are based on the leave-one-out cross-validation protocol. Due to the fact that the training set in dissimilarity representations

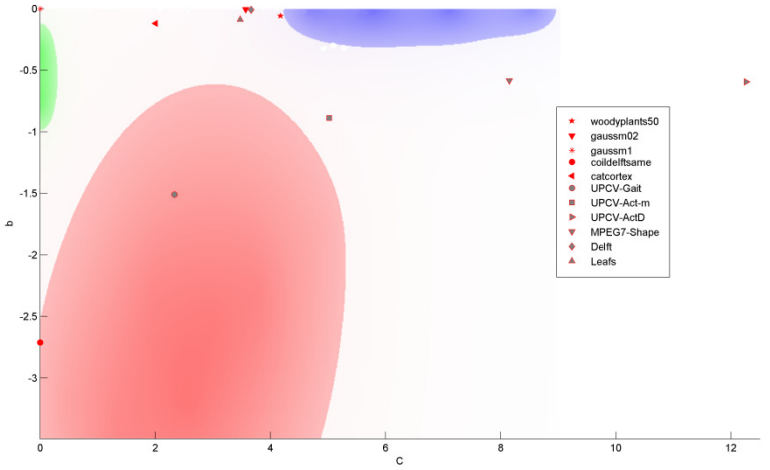
consists of the dissimilarity matrix of the training samples, the dictionary used for CR is complete. As explained in the previous section though, in order to reinforce the sparsity of the obtained CR, especially in the case of SRC, the dictionary has to be overcomplete. To this purpose, after normalization we apply PCA to each dataset in order to reduce the dimensionality of the representation vectors. The projection matrix was computed in each run using the corresponding training set.

The obtained results are shown in Table 1. One can observe that SRC outperforms SVM in all datasets. In a first attempt to specify the characteristics of a dataset that favor CR-based methods for classification, one can observe that the advantage of CR-based classification compared to SVM is greater on datasets with smaller samples-per-class ratio such as the UPCV-Gait dataset. This can be explained due to the sparsity induced during CR via the l_1-norm regularization, which constitutes the representation stable even for very small representative sets. Concerning the behavior of CR-based classification regarding non-metric and non-Euclidean properties of the datasets, we cannot draw reliable conclusions from the above results, since SRC is able to perform better than SVM both on metric and non-metric datasets, as also in datasets within a wide range of NEF. These results raise the question whether exist appropriate markers, indicating than the incorporation of a CR-based method could be beneficial to the classification accuracy, comparing to the standard SVM.

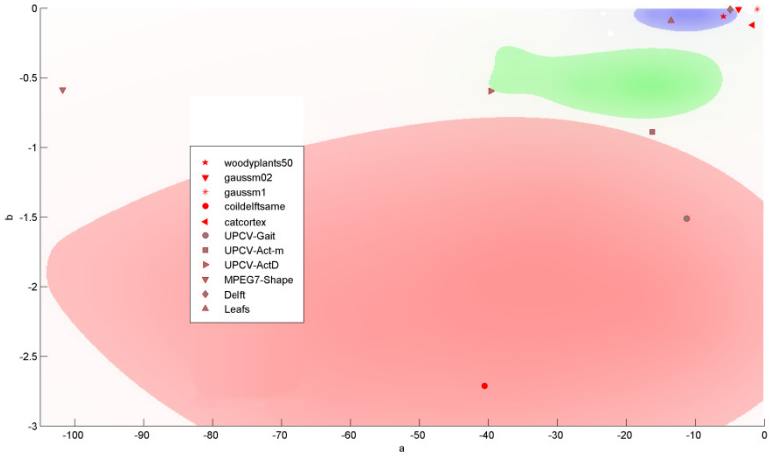
In order to model structural characteristics of the utilized datasets, thus forming appropriate criteria for the utilization of CR-based methods in the classification of dissimilarity data, we follow the rationale of Xu et al. in [31]. In their work, they considered three sources of non-Euclidean/non-Metric behavior in dissimilarity data: a) The manifold structure of the underlying data that implies non-Euclidean behavior to the derived dissimilarities, b) The spatially extended nature of the data samples forcing distance to be measured between the closest points of the surface of two samples, resulting pairwise distances that violate the triangle inequality, and c) Additive Gaussian noise to the originally Euclidean dissimilarities, resulting both non-Euclidean and non-metric data. In the same work authors introduced some empirical measures in order to identify the source of non-metric and non-Euclidean behavior of an arbitrary dataset. To this purpose, they proposed to model the negative spectrum of the Gram matrix corresponding to a dissimilarity matrix, by fitting a simple exponential function of the form $y(x) = a \cdot e^{bx}$, with b being the slope and a the intercept. Furthermore, as a measure for the characterization of the non-metric behavior, the parameter $C = \max_{i,j,k} \left| d_{ij} + d_{ik} - d_{jk} \right|$ is computed, where d_{ij} is the dissimilarity between the i^{th} and j^{th} data sample. The parameters *a*,*b* and *C* are used as measures, characterizing the negative spectrum and therefore the whole dataset.

We calculated the above parameters for all the utilized datasets and the results are illustrated in figure 1. In figure 1.a the datasets are distributed according to the corresponding values of parameters *b* and *C*. Figure 1.b illustrates the distribution of datasets according to the values of *a* and *C*. The area marked in green on both plots, indicate the space where the datasets characterized by manifold structure of their data lie in. The area marked in red, indicate the domain of datasets characterized by the spatially extended nature of their data samples. Finally, the area marked in blue

indicates the domain of the datasets, in which Gaussian noise has been added to the initially Euclidean dissimilarities. The illustration is based on the findings in [31], where artificially data from the three categories were used in order to define the corresponding regions. Apparently, the unmarked area corresponds to intermediate or mixed states, where more than one source of non-metric/mom-Euclidean behavior is inherent in the data.



(a)



(b)

Fig. 1. The distribution of the datasets used for evaluation, according to (a) The slope b and the metric constant C; (b) The slope b and the intercept a. The area marked in green corresponds to datasets that exhibit non-Euclidean behavior due to manifold structure of their data. The area marked in red, indicate datasets with spatially extended objects, and the area in blue indicate additive Gaussian noise to initially Euclidean dissimilarities.

As illustrated in Figure 1, the most of *UPCV-Dissim* datasets are characterized by mixed non-Euclidean/ non-metric behavior, where more than one sourse of negative spectrum and triangular violations affect the data. Specifically, *UPCV-Action-m* dataset is likely to contain extended objects lying and a loose manifold structure of the data. *Delft* and *Leafs* datasets are expressing a manifold behavior mixed with noise. The *UPCV-Gait* dataset probably contains extended objects in the underlying data samples, causing the triangular violations. Finally, we cannot be conclusive about the source of the properties of *MPEG7-Shape* and *UPCV-ActionD* datasets, although the large percentage of triangular violations and high NEF indicate the presence of extended objects on a manifold data structure, possibly with some additive noise.

Regarding the properties that favor the CR-based methods, it is easily inferred that such schemes perform significantly better than SVM on datasets characterized by the presence of spatially extended objects. Also, in the intermediate states which are characterized by low noise, CR-based methods seem to have a clear advantage. Things become vague as the properties of a dataset resemble the noisy case. There are many cases though, that due to limited number of available samples for training, CR-based methods proved to be advantageous in noisy conditions, such as *Delft*, *Leafs* and *Woodyplants50* datasets.

5 Conclusions

In this work we performed a thorough evaluation of the collaborative sparse representation SRC scheme, on the classification task in dissimilarity space. In particular we examined the performance of the most popular CR-based classification scheme, using a large set of public domain dissimilarity datasets, representing a wide range of tasks requiring classification of visual information. We compared the performance of the above methods to that of linear SVM scheme which is considered to be a landmark to the classification of dissimilarity data. We showed that CR-based classification can offer a clear advantage in challenging situations characterized by extreme non-metric and non-Euclidean behavior, as well as limited number of available training samples per class. Furthermore, we investigated the structural qualities of a dataset that constitute the CR-based classification beneficial compared to the SVM. To this purpose we utilized a three-parameter modeling of the non-metric and non-Euclidean properties of a dataset, corresponding to three independent sources of such behavior.

We demonstrated that CR-based methods outperform SVM on classifying dissimilarity data that contain spatially extended objects, manifold structure of the underlying data or a combination of these qualities, even if a small amount of noise has infiltrated to the data. We also showed that the induced sparsity during CR in the SRC scheme is of great significance to the classification performance, especially in cases with small representative sets in the training data and large number of classes.

Acknowledgements. This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

References

1. Bellman, R.: Dynamic Programming. Princeton University Press (1957)
2. Bunke, H., Buhler, U.: Applications of approximate string matching to 2D shape recognition. Pattern Recognition 26(12), 1797–1812 (1993)
3. Bunke, H.: On a relation between graph edit distance and maximum common subgraph. Pattern Recognition Letters 18(8), 689–694 (1997)
4. Cox, T.F., Cox, M.A.A.: Multidimensional Scaling. Chapman & Hall, London (2001)
5. Duin, R.P.W., Pękalska, E.: Non-Euclidean Dissimilarities: Causes and Informativeness. In: Hancock, E.R., Wilson, R.C., Windeatt, T., Ulusoy, I., Escolano, F. (eds.) SSPR & SPR 2010. LNCS, vol. 6218, pp. 324–333. Springer, Heidelberg (2010)
6. Duin, R.P.W., Loog, M., Pękalska, E.z., Tax, D.M.J.: Feature-Based Dissimilarity Space Classification. In: Ünay, D., Çataltepe, Z., Aksoy, S. (eds.) ICPR 2010. LNCS, vol. 6388, pp. 46–55. Springer, Heidelberg (2010)
7. Elhamifar, E., Sapiro, G., et al.: Finding Exemplars from Pairwise Dissimilarities via Simultaneous Sparse Recovery. In: Advances in Neural Information Processing Systems (2012)
8. Fang, S.C., Chan, H.L.: Human Identification by quantifying similarity and dissimilarity in electro-cardiogram phase space. Pattern Recognition 42, 1824–1831 (2009)
9. Fotopoulou, F., Theodorakopoulos, I., Economou, G.: Fusion in Phase Space for Shape Retrieval. In: EUSIPCO (2011)
10. Fotopoulou, F., Laskaris, N., Economou, G., Fotopoulos, S.: Advanced Leaf Image Retrieval via Multidimensional Embedding Sequence Similarity (MESS) Method. Pattern Analysis and Applications (2011)
11. Friedman, J.H., Rafsky, L.C.: Multivariate Generalizations of the Wald-Wolfowitz and Smirnov Two-Sample Tests. Annals of Statistics 7(4), 697–717 (1979)
12. Goldfarb, L.: A unified approach to pattern recognition. Pattern Recognition 17, 575–582 (1984)
13. Graepel, T., Herbrich, R., Bollmann-Sdorra, P., Obermayer, K.: Classification on pairwise proximity data. Advances in Neural Information System Processing 11, 438–444 (1999)
14. Graepel, T., Herbrich, R., Schölkopf, B., Smola, A., Bartlett, P., Muller, K.R., Obermayer, K., Williamson, R.: Classification on proximity data with LP-machines. In: ICANN, pp. 304–309 (1999)
15. Hammer, B., Mokbel, B., Schleif, F.-M., Zhu, X.: Prototype-Based Classification of Dissimilarity Data. In: Gama, J., Bradley, E., Hollmén, J. (eds.) IDA 2011. LNCS, vol. 7014, pp. 185–197. Springer, Heidelberg (2011)
16. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the Hausdorff distance. PAMI 15(9), 850–863 (1993)
17. Jacobs, D., Weinshall, D., Gdalyahu, Y.: Classification with Non-Metric Distances: Image Retrieval and Class Representation. IEEE TPAMI 22(6), 583–600 (2000)
18. Laub, J., Roth, V., Buhmannb, J.M., Müller, K.R.: On the information and representation of non-Euclidean pairwise data. Pattern Recognition, 1815–182639 (2006)
19. Lichtenauer, J., Hendriks, E.A., Reinders, M.J.T.: Sign Language Recognition by Combining Statistical DTW and Independent Classification. PAMI 30, 2040–2046 (2008)
20. Pekalska, E., Duin, R.P.W., Paclik, P.: Prototype selection for dissimilarity-based classifiers. Pattern Recognition 39(2), 189–208 (2006)
21. Pekalska, E., Duin, R.P.W.: The dissimilarity representation for pattern recognition. World Scientific (2005)

22. Plasencia-Calaña, Y., Orozco-Alzate, M., et al.: Selecting feature lines in generalized dissimilarity representations for pattern recognition. Digital Signal Processing (2012)
23. Schleif, F.-M., Zhu, X., Hammer, B.: A Conformal Classifier for Dissimilarity Data. AIAI (2), 234–243 (2012)
24. Schölkopf, B., Smola, A., Müller, K.R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10, 1299–1319 (1998)
25. Castrodad, A., Sapiro, G.: Sparse Modeling of Human Actions from Motion Imagery. Int. J. Comput. Vision 100, 1–15 (2012)
26. Theodorakopoulos, I., Kastaniotis, D., Economou, G., Fotopoulos, S.: Pose-based Human Action Recognition via Sparse Representation in Dissimilarity Space. J. Vis. Commun. Image R (2013)
27. Torgerson, W.S.: Multidimensional scaling: I. Theory and method. Psychometrika 17, 401–419 (1952)
28. Wilson, R., Hancock, E.: Spherical embedding and classification. In: SSPR (2010)
29. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust Face Recognition via Sparse Representation. PAMI 31, 210–227 (2009)
30. Xu, W., Wilson, R.C., Hancock, E.R.: Determining the Cause of Negative Dissimilarity Eigenvalues. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds.) CAIP 2011, Part I. LNCS, vol. 6854, pp. 589–597. Springer, Heidelberg (2011)
31. Xu, W., Wilson, R.C., Hancock, E.R.: Determining the Cause of Negative Dissimilarity Ei-genvalues. In: CAIP (2011)
32. Zhag, L., Yang, M., Feng, X.: Sparse representation or collaborative representation: Which helps face recognition? In: ICCV (2011)