Multiple Images Set Classification via Network Modularity

Thiago H. Cupertino[1], Thiago C. Silva, Liang Zhao
Departamento de Ciências da Computação
Instituto de Ciências Matemáticas e de Computação - ICMC
Universidade de São Paulo - Campus de São Carlos - Caixa Postal 668
13560-970 São Carlos, SP, Brazil.
{thiagohc, thiagoch, zhao}@icmc.usp.br

***Abstract*—In this paper, we consider image sets classification problem. For each image, a set of observations is obtained, which consists of different transformations, possibly including rotation, occlusion, projection and distortion, of the same image pattern. Each set belongs to a single pattern, i.e., the pattern is considered invariant under such transformations. The method employs a network representation of the input data to take advantage of the topological relations between the image patterns revealed by a low-dimensional manifold. In the constructed network, a measurement called modularity is computed to indicate the topological characteristics of the image network. Simulations are carried out in real image data sets and the results have showed that the proposed method outperforms some recent and state-of-the-art techniques.**

***Keywords*-multiple observations; image sets; modularity measure; classification; complex networks; pattern classification.**

I. Introduction

The amount of information stored and processed by computers has been growing fast. Different data types, such as financial transactions, image and sound databases, customer relations and scientific measurements have been handled. In this context, the last 20 years have witnessed a considerable growth in the study of automatic data processing algorithms, such as classification, clustering, information retrieval, association rules, among others [1], [2]. These algorithms are intended to automatically extract relevant information from the database to be explored.

In particular, multiple observations of the same object may be produced and stored in a database, as is the case, for example, of sensor networks. In these scenarios, the images of an object are captured from different viewpoints for further analysis [3], [4]. A car being tracked via a road camera system or a person having his/her images recorded and stored by an internal vision sensors network are examples of objects that are captured at different time instants or by different angles and geometric transformations. Thus, efficient methods must be developed to exploit multiple views of invariant patterns in order to correctly predict and extract relevant information, as is done in classification tasks.

Among the many machine learning methods developed for classification, the network-based learning algorithms have been explored with great success [5], [6]. Usually, these algorithms consist of two stages: the construction of a network from the original vector-based data set and the execution of an algorithm on the constructed network.

Another research area that has been received attention in recent years are the complex networks. A complex network refers to a large scale graph with nontrivial connection patterns [7], [8]. Such networks have emerged as a unified representation of complex systems in various branches of science [9]. Many network measures have been developed to characterize vertices, subnetworks and the whole network [10], [11]. Such measures can also be used to distinguish one type of network from another. One of the salient features of complex networks, as data representation tools, is the ability to describe the topological structure of the data. Such representation not only does emphasize the physical distances among vertices, but it also captures from local to global relations among the data. Consequently, it is a suitable tool to uncover pattern formation among data.

This paper focuses on the development of a a supervised technique where each pattern is composed of a group of representations instead of a single image. Each object is represented by a group of transformations and the classifier must predict the class this group of multiple images belongs to. The prediction is based on a low-complexity network measurement called modularity [12], [13]. Despite the simplicity of the method, it has been shown very competitive comparing to some recent and state-of-the-art methods.

The remainder of this paper is organized as follows. Section II introduces and formalizes the problem explored. Section III explains each step of the proposed method. In Section IV, some simulation results are provided and discussed and, finally, Section V concludes the paper.

II. The Problem of Multiple Observation Sets

In this section the problem of multiple observation sets classification is formalized as follows. Consider a pattern p has m multiple observations of the following form:

$$x_i{}^{(u)} = o_i(p),\ i = 1, \ldots, m, \tag{1}$$

where o_i represents different views of pattern p, such as rotation, scaling, perspective, projection etc. Superscript (u)

[1]This work is related to the author's ongoing doctoral thesis.

denotes that the set of observations is *unlabelled*, i.e., the set the pattern belongs to is unknown. The task is to classify p into one of the c classes considered in the problem by using the unlabelled observation set and labelled sets $X^{(l)}$.

Problem 1: Given labelled data sets $X^{(l)}$ and an unlabelled data set $X^{(u)} = \{x_j^{(u)} = o_j(p),\ j = 1, \ldots, m\}$ corresponding to multiple observations of p, the task is to correctly predict class c^* of the original pattern p, by using the group of representations $X^{(u)}$ related to pattern p.

III. Proposed Method

The method introduced in this paper consists of two main steps: network construction and modularity calculation. The classification task is then performed as the final step. The step of network construction is responsible for mapping the original image patterns to a low-dimensinal map relying on a network structure, where links represent the similarities between two neighbor patterns. The patterns are represented by the network nodes. Next, the modularity calculation step is performed and the classification task is based on its results. The modularity measurement is capable of quantifying, based on a network scheme, the similarity between the set of representations of the original patterns and the set of pattern transformations. Therefore, using this measurement, the class prediction is performed taking into account the set of transformations which is better related to the original pattern. The method is explained more detailed in the next subsections.

A. Network Construction

Networks are constructed from data sets by using the *nearest neighbors* (k-NN) method. This method consists in creating an edge between a vertex i and its k closest neighbors. In the original pattern set scenario, it is equivalent to finding the k most similar patterns of a reference pattern. Consider the labelled data set $X^{(l)}$ of a specific pattern p. First, the similarities among all patterns must be calculated, for instance, using a Gaussian function of the following form:

$$S_{ij} = exp\left(-\frac{||p_i - p_j||^2}{2\sigma^2}\right), \qquad (2)$$

where S_{ij} is the similarity between patterns i and j, $||.||$ is a norm measurement, and σ a width parameter.

As an example, Fig. 1 shows a pattern linked with its two nearest neighbors. In a similarity network, if a pattern i has pattern j as its nearest neighbor, the reciprocal is not true, i.e., pattern j can have pattern l as its nearest neighbor. Therefore, in this work the edges are directed, representing only one direction of similarity.

The constructed network is a set $\{V, E\}$ composed of a set of vertices $V = \{v_i\}$ representing each pattern i and a set of edges $E = \{e_{ij}\}$. If there is an edge between patterns i and j, then $e_{ij} = 1$. If there is no edge, then $e_{ij} = 0$.

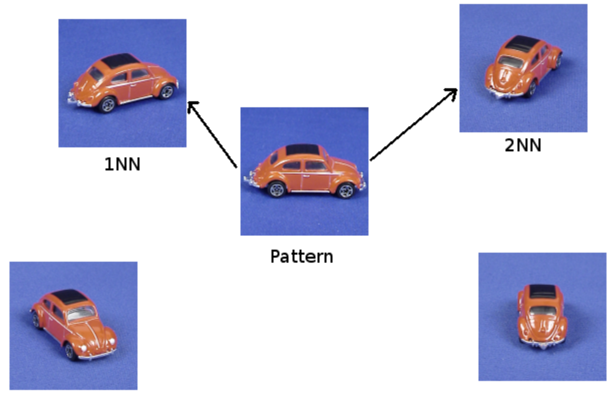


Fig. 1. An example of a k-NN network formation. The central pattern is linked with its two nearest neighbors: 1NN and 2NN.

B. Modularity Calculation

Consider a network with g densely connected subgroups and the connections among these groups are sparse. The modularity measurement [12] is defined as

$$Q = \sum_i (c_{ii} - {a_i}^2) = Tr(C) - ||C^2||, \qquad (3)$$

where C is a symmetric matrix whose element c_{ij} is the fraction of all edges in the network that link vertices in groups i and j, $||X||$ is the sum of elements of matrix X, $Tr(X)$ is the trace of matrix X, and a_i is defined as the column (or row) sum: $a_i = \sum_j c_{ij}$. The modularity measures the fraction of the edges in the network that connect vertices within the same densely connected subgroup minus the expected value of the same quantity in a network with the same group divisions, but considering random connections among vertices.

The trace of matrix C gives the fraction of edges in the network that connect vertices in the same dense group. A good division of the network should result in a high value of this trace. Variable a_i gives information about the fraction of edges that connect the vertices in group i. Therefore, if the number of connections inside a group is no better than random, $Q = 0$. When Q approaches 1, it means that the network has a strong division considering the g identified dense subgroups.

C. Classification of Multiple Observation Sets via Modularity Measurement

The method for the classification of multiple pattern observations deals with the problem defined in Problem 1, Section II. First, the method constructs a k-NN network to represent the topological relations among the labelled patterns $X^{(l)}$. Each pattern set is represented by one exclusive network. Then, if we are given P different pattern groups, the method constructs P different networks, each one corresponding to a different pattern class. This step can be viewed as the training phase.

Next, the method computes the modularity measurement associated with the unlabelled sets $X^{(u)}$. To do so, the unlabelled set $X^{(u)}$ is joined to each constructed network using the same

k-NN network formation method: each unlabelled pattern in $X^{(u)}$ is connected to its k most similar neighbors. These similar neighbors can be selected from the constructed network as well as from the unlabelled set. After joining the unlabelled set $X^{(u)}$, a new network is formed. Then, the modularity measurement Q is calculated according to Eq. 3 considering $g = 2$ densely connected subgroups, i.e., one is the group of densely connected vertices corresponding to the initial constructed k-NN network and the other group corresponds to the unlabelled vertices joined to the constructed network. This process is performed for the P different networks constructed in the training phase. At the end of the process, the unlabelled group is classified as belonging to the pattern network whose Q has presented the lowest value, indicating that the unlabelled patterns are more similar to the patterns in that network.

Algorithm

In a concise form, the proposed method can be summarized by Algorithm 1.

Algorithm 1 : Proposed Modularity Method

Input:
$X^{(l)}$: labelled data sets
$X^{(u)}$: unlabelled set
c : number of classes
l : number of labelled sets
Parameters:
k : number of nearest neighbors
Output:
c^* : Estimated class for the unlabelled data
Training:
for $i = 1 \to l$ **do**
  $Net_i^{(l)}$ = Create k-NN network from labelled set $X_i^{(l)}$
end for
Classification:
for $i = 1 \to l$ **do**
  Net_i = Join unlabelled set $X^{(u)}$ into $Net_i^{(l)}$
  Q_i = Compute modularity from Net_i
end for
$c^* = argmin_i(Q_i)$

Computational Complexity

The computational complexity of Algorithm 1 is determined mainly by the k-NN network construction in the training phase, as the modularity calculation can be approximated by a logarithm complexity [13]. The network construction takes $O(n^2)$, where n is the number of labelled patterns, due to the dissimilarity calculation among all patterns. In the classification phase, each constructed network takes $O(n*m)$, where m is the number of unlabelled observations os the same patter. Therefore, considering both phases, the algorithm runs in $O(n^2)+O(n*m)$. On the other hand, as the training phase must run only once before usage, the method takes $O(n*m)$ to classify a new set of multiple observations.

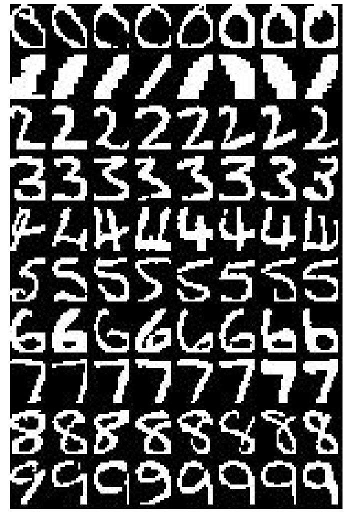


Fig. 2. Samples from the digit images data set. Each class, from "0" to "9", corresponds to the same invariant pattern.

IV. RESULTS

In this section the proposed modularity method is tested on two real databases publicly available. One is composed of handwritten digits collections and the second is composed of multi-view images. The method is also compared to some state-of-the-art algorithms.

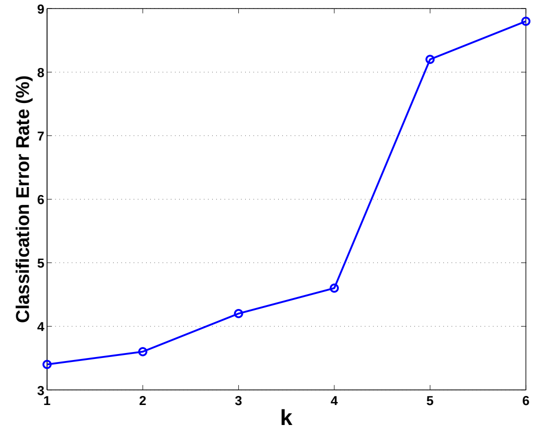
A. Handwritten Digits

The first simulations were performed on a digit images database [14]. This collection contains 20×16 pixels binary images from "0" to "9". Each of the 10 classes contains 39 examples. Some samples can be seen in Fig. 2. The simulations in this subsection had the following configuration: first, the data sets were randomly split into training and test sets. The training sets were composed of two random examples per class. Each example was augmented by four new virtual examples generated by successive rotations of the original example. The rotation angles were sampled regularly in $[-40°, 40°]$. The test sets were composed of m (Eq. 1) randomly chosen examples and each example was rotated by a randomly chosen angle in the same interval $[-40°, 40°]$, in a uniform manner. These virtual samples were generated to account for the robustness to pattern transformations. The use of these samples can reinforce the transformation invariance into classification algorithms [15]. Therefore, the classification method becomes more robust to transformations of the test instances. Values of $m = 10, 30, 50, 70, 90, 110, 130$ and 150 were considered for the test set size.

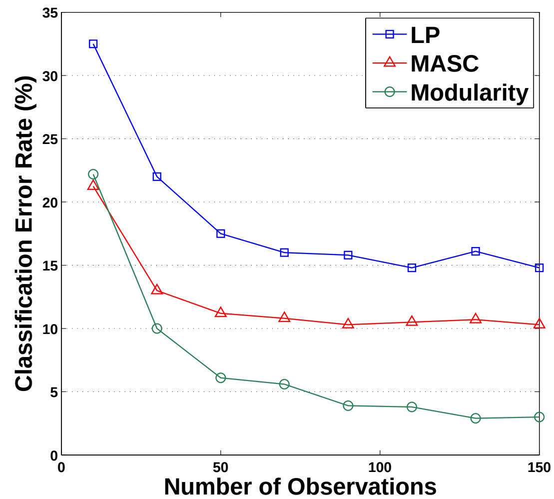
We compared the proposed modularity method with two other network-based methods: Label Propagation (LP) [16] and Manifold-Based Smoothing under Constraints (MASC) [15]. The LP algorithm forms a k-NN network with weighted edges. Next, it computes a real-valued label matrix via a regularization framework function with a cost function [16]. The idea is to find a label matrix which is smooth along the edges of similar pairs of vertices and at the same time close to the initial labels (the labelled set $X^{(l)}$). On the other hand, the MASC algorithm is a specialized version of the LP algoritm to deal with the problem of multiple observation sets. Since all test samples belong to the same class in this problem, the optimal solution can be obtained with a full search, as long as the number of classes stays reasonable. Thus, MASC has been formulated as a discrete optimization problem [15].

As the modularity method uses the k-NN network formation method, the results for different values of k were analysed. In all cases, the dissimilarity among the examples was calculated using Eq. 2 with a unit standard deviation (by setting $\sigma = 1/\sqrt{2}$). Fig. 3a shows that the higher the value for parameter k, the higher the classification error. In other words, the modularity method must use low values for k to achieve the best results.

The results are shown in Fig. 3b. For each value of m ranging from 10 to 150, 1000 runs were averaged, corresponding to 100 runs for each of the 10 classes. For all algorithms, we set $k = 5$, the value for which the LP algorithm achieved its best results. Despite its simplicity, the modularity method achieved the best results for most test set numbers of observations m. Furthermore, the results were even better when k took lower values (Fig. 3a).



(a)



(b)

Fig. 3. (a) Classification error rates (%) using different values for parameter k on the handwritten digits images data set. Each value was averaged over 50 runs. The number of observations for the test set is $m = 30$. (b) Simulation results for the digit images data set. All algorithms used $k = 5$ for the network construction. Each point was averaged over 1000 runs.

B. Image Collections

The next data set simulated was an image collection of multiple views of objects called ETH-80 [17]. All patterns of this set are shown in Fig. 4a. In total, there are 80 different classes, each one composed of 41 different views, totalling 3280 instances. Each class belongs to one of the following categories: apple, pear, tomato, cow, dog, horse, cup and car. As an example, the class "red car" can be viewed in Fig. 4b. It must be stressed that the invariant patterns in this data set suffer from different rotation angles.

In order to compare the proposed modularity method, four methods specialized in multiple image observations were simulated: Manifold-Based Smoothing under Constraints (MASC) [15], summarized in the previous subsection, Mutual Subspace Method (MSM) [18], Kernel Mutual Subspace Method (KMSM) [19] and KL-divergence (KLD) [20]. In a few words, these methods work as follows:

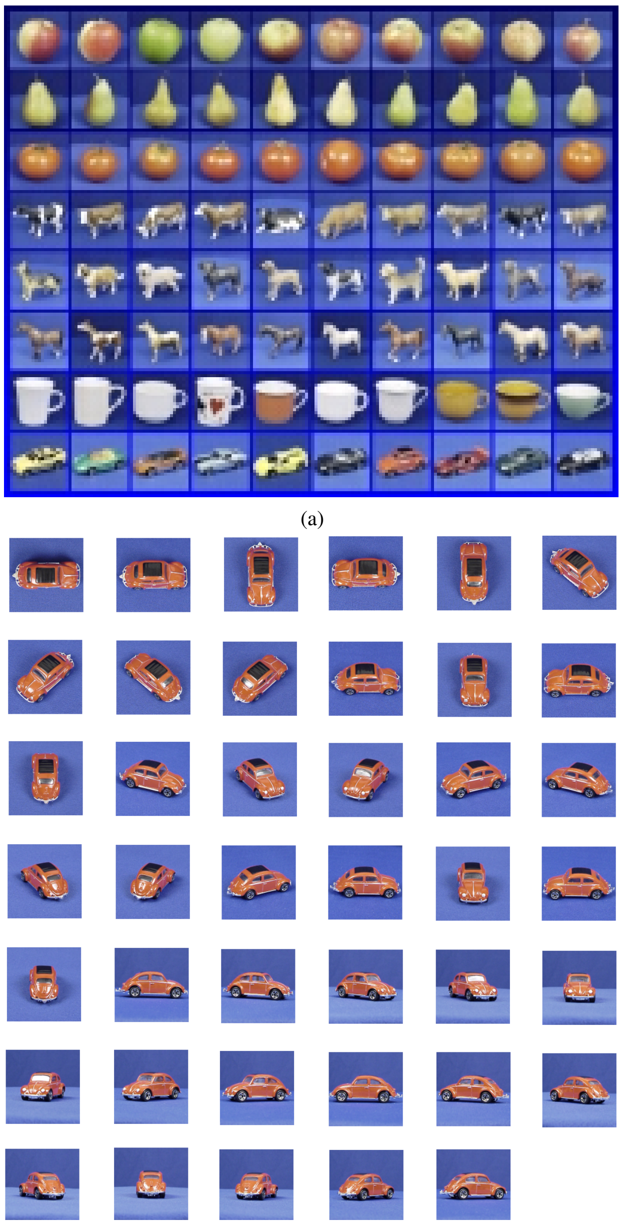
- MSM: It represents each image by its principal components and computes the similarity between a test and a training image via principal angles [21].
- KMSM: It is an extension of the MSM method. Instead of using the PCA to compute the image principal component, it uses kernel PCA to take into account nonlinearities.
- KLD: It formulates the problem of classification of multiple observations of images as a statistical hypothesis test. Each set is assumed to fit a Gaussian distribution and the method computes the KL-divergence among the sets.

The simulations were performed with the following configurations: the original image size is 128×128 pixels but, for computations ease, the images were downsampled to 32×32; the training and test sets were composed of 21 and 20 samples randomly chosen from the original sets, without repetition; the dissimilarity between patterns was calculated using Eq. 2 with a unit standard deviation (by setting $\sigma = 1/\sqrt{2}$), after extracting and comparing the patterns' features. These features were extracted by the spatiogram measurement [22] and compared using the discrete Bhattacharyya coefficient [23] (as the norm measurement in Eq. 2). Spatiograms are able to capture higher-order spatial moments. A 2^{nd}-order spatiogram model of an object is identical to a histogram of its features, except that it also stores additional spatial information, namely the mean and covariance of the spatial position of all pixels that fall into each histogram bin.

In order to compare the results, the value of $k = 5$ was used in the simulations (the value for which the LP algorithm achieved its best results), as in the previous subsection - although for the proposed method the results were even better

TABLE I
CLASSIFICATION ACCURACY AVERAGE (%) (STANDARD DEVIATION) FOR THE ETH-80 IMAGES DATA SET.

MASC	MSM	KMSM	KLD	Proposed Modularity Method
88.88 (1.71)	74.88 (5.02)	83.25 (3.40)	52.50 (3.95)	**92.71(2.65)**



(a)

(b)

Fig. 4. (a) The 80 different patterns in ETH-80 data set. (b) The 41 different views composing the "red car" class in ETH-80 data set.

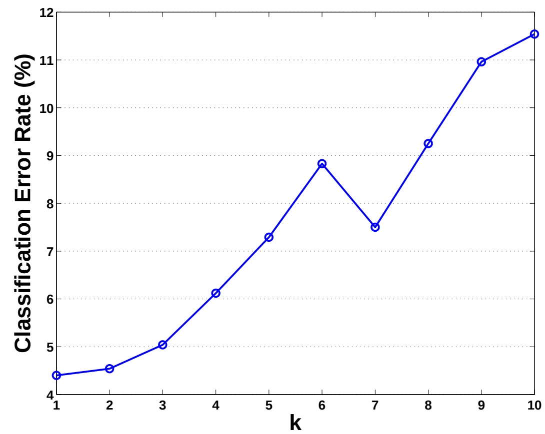


Fig. 5. (a) Classification error rates (%) using different values for parameter k on the ETH-80 data set. Each value was averaged over 50 runs. The number of observations for the test set is $m = 30$.

using smaller values for k, as it can be seen on Fig. 5. Table I shows that the proposed modularity method achieved the best classification results (avereged over 50 runs) for the ETH-80 data set, overcoming an already good result obtained by the MASC method.

V. CONCLUSIONS

A new network-based method for the classification of multiple observation sets has been proposed in this paper. In the context studied, each multiple observation set correponds to a single pattern. The method consists of two stages: network construction and modularity calculation. The network construction provided a topological representation of the relations among the different patterns and the modularity measurement numerically represented the connectedness of the constructed network: the more the network was close to a single densely connected component, the more probable the objects could belong to the same pattern. The simulation results showed that the proposed method performed well in the handwritten digits and multiple views objects collections, overcoming many recent and state-of-the-art multi-view classification methods. As future extensions we suggest the study of different network measurements, which could take into account different characteristics of the topological representations.

This work is related to the author's ongoing doctoral thesis. During his doctoral research, he has presented his ongoing works in several conferences and has authored and co-authored 5 articles in international journals (see [24], [25], [26], [27], [28]).

ACKNOWLEDGMENTS

The authors would like to acknowledge the São Paulo State Research Foundation (FAPESP) and the Brazilian National Council for Scientific and Technological Development (CNPq) for the financial support.

REFERENCES

[1] O. Chapelle, B. Schölkopf, and A. Zien, Eds., *Semi-Supervised Learning*. Cambridge, MA: MIT Press, 2006. [Online]. Available: http://www.kyb.tuebingen.mpg.de/ssl-book

[2] X. Zhu, "Semi-supervised learning literature survey," Computer Sciences, University of Wisconsin-Madison, Tech. Rep. 1530, 2007.

[3] T. Zhao, R. Nevatia, and B. Wu, "Segmentation and tracking of multiple humans in crowded environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1198–1211, 2008.

[4] E. Kokiopoulou and P. Frossard, "Distributed classification of multiple observation sets by consensus," *IEEE Transactions on Signal Processing*, vol. 59, pp. 104–114, 2011.

[5] A. Micheli, "Neural network for graphs: A contextual constructive approach," *IEEE Transactions on Neural Networks*, vol. 20, pp. 498–511, 2009.

[6] H. Bunke and K. Riesen, "Recent advances in graph-based pattern recognition with applications in document analysis," *Pattern Recognition*, vol. 44, pp. 1057–1067, 2011.

[7] R. Albert and A.-L. Barabsi, "Statistical mechanics of complex networks," *Review of Modern Physics*, vol. 74, pp. 47–97, 2002.

[8] M. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[9] S. Bornholdt and H. G. Schuster, *Handbook of Graphs and Networks: From the Genome to the Internet*, 1st ed. Wiley-VCH, 2003.

[10] L. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas, "Characterization of complex networks: A survey of measurements," *Advances in Physics*, vol. 56, no. 1, pp. 167–242, 2007.

[11] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics Reports-review Section of Physics Letters*, vol. 424, pp. 175–308, 2006.

[12] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks." *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 69, no. 2, 2004.

[13] B. DasGupta and D. Desai, "On the Complexity of Newman's Community Finding Approach for Biological and Social Networks," *Computing Research Repository*, vol. abs/1102.0, 2011.

[14] S. Roweis. (2012) New York University. [Online]. Available: http://www.cs.nyu.edu/ roweis/data.html

[15] E. Kokiopoulou and P. Frossard, "Graph-based classification of multiple observation sets," *Pattern Recognition*, vol. 43, pp. 3988–3997, 2010.

[16] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schlkopf, "Learning with local and global consistency," in *Neural Information Processing Systems*, 2003.

[17] B. Leibe and B. Schiele, "Analyzing appearance and contour based methods for object categorization," in *Computer Vision and Pattern Recognition*, 2003, pp. 409–415.

[18] K. Fukui and O. Yamaguchi, *Face Recognition Using Multi-viewpoint Patterns for Robot Vision*, 2003.

[19] H. Sakano and N. Mukawa, "Kernel mutual subspace method for robust facial image recognition," in *Knowledge-Based Intelligent Information and Engineering Systems*, 2000, pp. 245–248.

[20] G. Shakhnarovich, J. W. F. III, and T. Darrell, *Face Recognition from Long-Term Observations*, 2002.

[21] A. Hadid and M. Pietikinen, "From still image to video-based face recognition: An experimental analysis," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 813–818.

[22] O. Conaire, N. E. O'Connor, and A. F. Smeaton, "An improved spatiogram similarity measure for robust object localisation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2007, pp. 1069–1072.

[23] A. Djouadi, O. Snorrason, and F. D. Garber, "The quality of training sample estimates of the bhattacharyya coefficient," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 92–97, 1990.

[24] T. H. Cupertino, T. C. Silva, and L. Zhao, "Classification of multiple observation sets via network modularity," *Neural Computing and Applications*, Accepted for publication.

[25] T. H. Cupertino, J. Huertas, and L. Zhao, "Data clustering using controlled consensus in complex networks," *Neurocomputing*, In submission.

[26] T. C. Silva, L. Zhao, and T. H. Cupertino, "Handwritten data clustering using agents competition in networks," *Journal of Mathematical Imaging and Vision*, Accepted for publication.

[27] R. Estepa, A. Estepa, T. H. Cupertino, J. Vozmediano, and G. Madinabeitia, "A productivity-based approach to lan topology design," *Communications Letters, IEEE*, vol. 15, no. 3, pp. 349 –351, march 2011.

[28] R. Estepa, A. Estepa, and T. H. Cupertino, "A productivity-oriented methodology for local area network design in industrial environments," *Computer Networks*, vol. 55, no. 9, pp. 2303 – 2314, 2011.