Dissimilarity-based Representation for Local parts

A. Carli¹, U. Castellani¹, M. Bicego¹,², V. Murino¹,²

¹Department of Computer Science, University of Verona, Strada Le Grazie 15, 37134 Verona (Italy)  
²Istituto Italiano di Tecnologia (IIT), Via Morego 30, 16163 Genova (Italy)

In recent years, the dissimilarity-based representation paradigm has aroused a lively interest in the pattern recognition community. This paradigm differs from typical pattern recognition approaches where objects to be classified are represented by feature vectors. In the dissimilarity-based paradigm, objects are described using pairwise (dis)similarities. In this way, objects are not constrained to be explicitly represented in a feature space, and all that is necessary is a way to compute (dis)similarities between pairs of objects. The goal is then to learn a classifier only from these relational data. The general idea is the following: given a set of pairwise dissimilarity values, a new representation space can be built, in which each object is described by these values. In the original versions, a given object was characterized with dissimilarities/similarities from other objects in the data set, the so-called prototypes -- obviously the choice of the prototypes represents a crucial issue. This idea has been then refined and generalized, considering also dissimilarities from class-models, cluster models, or even components of a model.

In this paper, a novel contribution in this direction is proposed, considering and exploiting dissimilarities between parts of a given object. The basic idea consists of defining the set of prototypes in terms of relevant object subparts extracted from the training set. Part-based characterization of patterns has been largely applied in many computer vision applications, e.g., object categorization and image retrieval, where an image is decomposed into parts (or local features). Here, we focus on the object classification problem in which the objects are images described in term of local parts.

The typical approach to extract image parts is to employ an interest region detector. Such parts are then characterized using a proper descriptor (like SIFT, shape context, or others). The number of parts may vary from image to image and there is no ordering among features in a single image. As a result images are represented as variable-sized sets of unordered features, which makes not possible to apply standard vector-based classification algorithms. In the literature, this problem has been addressed by the family of the so-called Local Kernels. For instance a widely applied technique is the Bag of Keypoints (BoK). Being inspired by the Bag-of-Words approach for text classification, this method consists in transforming the set of features in a histogram, which counts the number occurrences in that image of a given set of visual words (i.e., prototype features). The Bag-of-Keypoints approach has been extended by combining the number of bins in a hierarchical fashion leading to the so called Pyramid-kernel-matching paradigm. Similarly, such hierarchical approach has been successfully exploited also in the spatial domain.

In this paper, we will approach the above mentioned problems using concepts and tools of the dissimilarity-based representation paradigm, showing that some of the Local Kernels (like BoK) are just special cases of our methodology. This may also open the new possibility of directly applying some of the results recently proposed in the general dissimilarity-based classification paradigm to the image classification problem.

Some experiments on a scene categorization task (using the ETH-80 dataset) show the effectiveness of the proposed approach, also in comparison to other Local Kernels present in the state of the art.