Computer Vision and Image Understanding 122 (2014) 155–171

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

On the use of supervised features for unsupervised image categorization: An evaluation ☆

Gianluigi Ciocca [a,1], Claudio Cusano [b,*,1], Simone Santini [c,1], Raimondo Schettini [a,1]

[a] Department of Informatics, Systems and Communication (DISCo), Università degli Studi di Milano-Bicocca, Viale Sarca 336, 20126 Milano, Italy
[b] Department of Electrical, Computer and Biomedical Engineering, Università degli Studi di Pavia, Via Ferrata 1, 2700 Pavia, Italy
[c] Escuela Politécnica Superior, Universidad Autónoma de Madrid, C/ Tomas y Valiente 11, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 28 March 2013
Accepted 28 January 2014
Available online 10 February 2014

Keywords:
Unsupervised image categorization
Supervised features
Primitive features
Image clustering

ABSTRACT

Recently, new high-level features have been proposed to describe the semantic content of images. These features, that we call supervised, are obtained by exploiting the information provided by an additional set of labeled images. Supervised features were successfully used in the context of image classification and retrieval, where they showed excellent results. In this paper, we will demonstrate that they can be effectively used also for unsupervised image categorization, that is, for grouping semantically similar images. We have experimented different state-of-the-art clustering algorithms on various standard data sets commonly used for supervised image classification evaluations. We have compared the results obtained by using four *supervised* features (namely, classemes, prosemantic features, object bank, and a feature obtained from a Canonical Correlation Analysis) against those obtained by using low-level features. The results show that supervised features exhibit a remarkable expressiveness which allows to effectively group images into the categories defined by the data sets' authors.

1. Introduction

Unsupervised categorization, often done through the use of clustering algorithms, is one of the most powerful techniques available to the designer of image management systems, as it allows categorization with no other information than that contained in the data themselves. Grouping images into semantically homogeneous classes is often a *sine qua non* for efficiently processing, structuring, querying, and browsing large collections of images. For instance, representative images can be extracted from each class to stand for the collection contents [1]; grouping similar images can also be useful for the design of effective user interfaces for browsing and visualization of image collections; image categories may be used to speed up database queries by pre-filtering the images to be searched [2], and so on. Alas, unsupervised categorization is also a very difficult problem. Without the information provided by class labels it is very difficult to obtain a reliable classification in semantically meaningful classes, and the performance of unsupervised classification is often nowhere near that of supervised methods. On the other hand, in applications one often faces the problem of categorizing a large, unstructured set of images not only without labeled training sets but, often, without *a priori* knowledge of the classes that are present in the collection.

Several authors have begun exploring features that, in addition to the image data, use semantic information in the guise of a set of labeled images belonging to a collection of pre-defined classes. These classes are not, in general, the same that we are interested in identifying in an unsupervised way, and the related labeled images come from a data set different from that which we are interested in classifying. In this paper we will consider specifically the work of Torresani et al. [3], Ciocca et al. [4], Li et al. [5] and Gordo et al. [6]. We shall refer to the features used in these papers as *supervised*, in a sense that will be clarified in the next section.

The purpose of this paper is to evaluate the performance of *supervised* features for unsupervised image categorization. First of all we verified if these features bring a significant improvement with respect to low-level features (which we shall call *primitive*). To this end we selected four data sets of different nature and four state-of-the-art clustering algorithms, and we compared the performance obtained by using supervised features with those obtained by using primitive features. We also verified how much the clustering performance depends on the dimensionality of the feature vectors. Finally, we determined whether the combination

☆ This paper has been recommended for acceptance by Nicu Sebe.

* Corresponding author.
E-mail addresses: ciocca@disco.unimib.it (G. Ciocca), claudio.cusano@unipv.it (C. Cusano), simone.santini@uam.es (S. Santini), schettini@disco.unimib.it (R. Schettini).

[1] The authors contributed equally to this work.

http://dx.doi.org/10.1016/j.cviu.2014.01.010
1077-3142/

of a simple clustering algorithm and supervised features could outperform other strategies, specifically designed for unsupervised image categorization. With these experiments we try to identify strengths and weaknesses of the different supervised features in dealing with different type of images.

In the last years, a huge amount of work and resources have been devoted to the evaluation of algorithms and systems for the supervised classification of images. This effort led to the collection of standard data sets and to the definition of experimental protocols culminating with the organization of public contests and challenges. The same cannot be told for the problem of unsupervised categorization. In this context, even though the focus of this paper is the evaluation of supervised features, we believe that it could also serve as a useful source of information about the performance of low-level state-of-the-art features.

The paper is organized as follows: Section 2 provides the definition of primitive and supervised features; presents a brief review of state-of-the-art high-level descriptors; and details the features included in the evaluation. Section 3 describes the four clustering algorithms considered. The experiments, including the performance measure, the data sets, and the results are reported in Section 4 and discussed in Section 5. Finally, Section 6 concludes the paper.

1.1. Related work

In the literature there are several works dealing with the problem of unsupervised image categorization that use either low-level or high-level features.

Among the works exploiting low level features we can cite Tuytelaars et al. [7]. In their works, a comparison of different clustering algorithms and different bag-of-words representations of scale invariant features are presented and tested by identifying ten categories extracted from the Caltech-256 data set, and the MSCR2 data set.

SIFT-like region descriptors, within a probabilistic latent semantic analysis, framework are used to discover objects categories in unlabeled images in [8]. Five object categories from the Caltech-101 (faces, motorbikes, airplanes, cars rear, and background) are used for experimentation.

Differently, Sivic et al. [9] try to automatically discover a semantically meaning hierarchical structure for images based on the visual appearance of objects. A visual vocabulary of quantized SIFT descriptors are used as image representation. Learning of the objects hierarchy is achieved using a generative hierarchical latent Dirichlet allocation. The hierarchy is used to recognize nine object classes (faces, cows, grass, trees, buildings, cars, airplanes, bicycles and sky).

The problem of scene category discovering is explicitly tackled in [10]. Different representations (Gist, SIFT, PACT and color) are used to describe the images content and an information projection strategy is used to identify informative and discriminative features. The scene categorization is treated as a graph partition problem and experiments are performed on the LHI eight scene categories and MIT eight scene categories data sets. A more recent work of the same authors [11] introduces the concept of weak training sets to be used for categorization learning. Different partitioning of the data set are learned using a max-margin classifier and these partitioning are combined into an ensemble proximity matrix which is fed to a spectral clustering algorithm.

In order to cope with the possible large variability within each image category some authors incorporate into the clustering process a local analysis of relevant parts in the images common to images belonging to the same category. Lee and Grauman [12] use a novel semi-local features to describe the images in terms of neighborhood appearance and geometry. Clustering is performed by an initial grouping based on feature correspondences and then it is iteratively refined based on the evolving intra-cluster pattern of local matches. Faktor et al. [13], introduced a similar approach named 'Clustering-by-Composition'. Categories are discovered by grouping images that share common statistically significant regions. These regions are those which have a low chance to occurring at random and are described in terms of HOG and Local Self-Similarity features.

Other recent studies have investigated unsupervised image categorization from a different perspective by exploring new clustering techniques and low-level descriptors. Käster et al. [14] tested k-means, Hierarchical Agglomerative Clustering, Partition Around Medoids and CLARA clustering algorithms on a subset of 1440 color images of 20 semantically disjoint object classes of the Columbia Object Image Library image collection. Images were described in terms of color moments, color distribution and structure. To evaluate the performance of the clustering algorithms with respect to semantically meaningful clusters the results were compared with a reference grouping by using the Rand-Index.

A spectral clustering algorithm named Locality Preserving Clustering has been presented by Zheng et al. [15]. The algorithm is based on a modified locality preserving projection algorithm and k-means clustering. The image descriptor is a 112 dimensional feature vector created by a combination of color histogram and color texture moments.

Grauman and Darrell [16] proposed a method where sets of local image features (SIFT descriptors compacted into ten-dimensional features via PCA) are compared in terms of partial match correspondences between component features, forming a graph between the examples that is partitioned via spectral clustering and normalized cut criterion.

Dueck and Frey [17] use affinity propagation to capture the underlying data structure. A non-metric similarity function based on SIFT features is used to group similar images belonging to a subset of 20 of the 101 classes in the Caltech101 data set.

The lack of semantic information provided by the class labels could be mitigated by using suitable high-level features. In this paper we will investigate whether or not those features, learned from labeled training sets, make it possible to achieve effective unsupervised image categorization.

Image labels can be in the form of textual keywords. For example, Loeff et al. [18] present a method exploiting a latent space induced by pre-annotated words associated to images. This intermediate feature space is created by using a max-margin factorization model that finds a low dimensional subspace with high discriminative power for correlated image annotations. A spectral clustering approach is finally applied to the representations in the latent space.

2. Supervised features

Several approaches have been investigated to automatically incorporate semantics into image representations [19]. Recently, features that use the semantic information provided by additional labeled images have been proved to be effective in a variety of image classification and retrieval tasks [20,3,5,21]. We argue that these features, which we call *supervised*, could perform well also in unsupervised image categorization.

2.1. Definition

Consider a reference database of images $D = \{x_1, \dots, x_n\}$ and a reference partition of D into classes (according to some semantically meaningful criterion), $\mathcal{D}$, with $\mathcal{D} = \{D_1, \dots, D_q\}, \bigcup_i D_i = D$, and for all $i \neq j, D_i \cap D_j = \emptyset$. The subsets D_i may or may not have associated labels.