Image and Vision Computing 27 (2009) 960–978

Contents lists available at ScienceDirect

Image and Vision Computing

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

Region and constellations based categorization of images with unsupervised graph learning

M.A. Lozano [a,*], F. Escolano [a], B. Bonev [a], P. Suau [a], W. Aguilar [b], J.M. Saez [a], M.A. Cazorla [a]

[a] Dpto. Ciencia de la Computación e Inteligencia Artificial, Universidad de Alicante, San Vicente del Raspeig, 03080 Alicante, Spain
[b] Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas, Universidad Nacional Autónoma de México, Apartado Postal 20-726, Ciudad Universitaria, C.P. 04510. México, D.F.

ARTICLE INFO

Article history:
Received 25 October 2007
Received in revised form 18 September 2008
Accepted 20 September 2008

Keywords:
Image categorization
Clustering of graphs
EM algorithms

ABSTRACT

In this paper, we address the problem of image categorization with a fast novel method based on the unsupervised clustering of graphs in the context of both region-based segmentation and the constellation approach to object recognition. Such method is an EM central clustering algorithm which builds prototypical graphs on the basis of either Softassign or fast matching with graph transformations. We present two realistic applications and their experimental results: categorization of image segmentations and visual localization. We compare our graph prototypes with the set median graphs. Our results reveal that, on the one hand, structure extracted from images improves appearance-based visual localization accuracy. On the other hand, we show that the cost of our central graph clustering algorithm is the cost of a pairwise algorithm. We also discuss how the method scales with an increasing amount of images. In addition, we address the scientific question of what are the bounds of structural learning for categorization. Our in-depth experiments both for region-based and feature-based image categorization, will show that such bounds depend hardly on structural variability.

1. Introduction

1.1. Categorization from region-based segmentation

Image categorization is a key problem in Computer Vision [1]. From the point of view of the efficiency of the image classification process, the prototype of a class yields a coarse-to-fine approach. In this regard, there are many examples relying on image segmentation as a previous step for categorization. For instance, in [2] it is argued that segmentation may provide interesting information about objects, even though the segmentation process is not perfect. Actually, many approaches have been addressed to a simplistic partitioning of the image in blocks (see for instance [3]). A well known example is Blobworld [4,5] where clustering/segmentation precedes image queries. In the latter approaches, few graph-based learning attempts have been performed in order to learn the structure of the adjacency-region graph and use it later for image indexing.

1.2. Categorization in the constellations approach

On the other hand, structural criteria, graph matching, and even graph learning, have been considered as fundamental elements in the set up of the constellation (part/features-based) approach to object recognition [23]. Most of the research in such direction has been focused on exploiting feature (local) statistics, whereas structural (global) statistics have been typically confined to the joint Gaussian of feature locations [13]. However, there has been recent interest in modelling and learning structural relationships. This is the case of the *tree-structured models* [14,21] and the *k-fans* graph model [10]. Anyway, models with higher relational power are often needed for solving realistic situations. In this regard, a key question is to find an adequate trade-off between the complexity of the model and the computational cost of learning and using it.

1.3. Central graph clustering

In this paper, we combine a recently developed method for graph clustering with a novel supergraph-based method for prototype building and also test experimentally such combination both in the region-based and in the constellation approaches for image categorization. Graph clustering is motivated by the need of different structural abstraction levels for categorizing images. One may follow either the central or the pairwise clustering approach. Here, we follow central graph clustering [9,25,16] because of its convenience for developing structural generative models, a task that has been more addressed in the context of trees [18,19] than in the more general case. However, the key of central graph clustering is to have an strategy for learning automatically the prototype of each class for further

* Corresponding author. Tel.: +34 965 90 38 97; fax: +34 965 90 39 02.
E-mail addresses: malozano@dccia.ua.es (M.A. Lozano), sco@dccia.ua.es (F. Escolano), boyan@dccia.ua.es (B. Bonev), pablo@dccia.ua.es (P. Suau), weam@turing.iimas.unam.mx (W. Aguilar), jmsaez@dccia.ua.es (J.M. Saez), miguel@dccia.ua.es (M.A. Cazorla).

0262-8856/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2008.09.011

analysis. Such task has been addressed by other researchers [18,17]. In [27], we proposed an incremental method which depends on the order in which the graphs are fused. Later, in [8], we presented an alternative method which overcomes such problem. It is based on the information provided by the diffusion kernels [11,22] in order to decide which matches are preferable to be considered in order to fuse the nodes of the graphs in the set. Our algorithm works both with continuous graph-matching methods like Softassign, or its kernelized version [26], and with faster alternative discrete matching methods, that is, no graph matching method is precluded. Thus, here we test both a structure-driven matching approach like Softassign and a highly attribute-based method like GTM, *graph-transformation matching* [6]. GTM relies in initial putative matches based on feature similarities (actually GTM works as an structural filter over the initial matches) yielding a *consensus graph*, provided that such subgraph exists.

1.4. Contributions and paper organization

Our contribution is twofold. First, we propose an efficient integration of graph clustering and supergraph-based prototype building; our graph-learning method is tested, using Softassing and GTM, in categorization of segmented images and in a *visual localization* (scene recognition) context. Besides this test, our second contribution is to determine what is the added value of structural categorization. In region-based categorization, this implies to analyze the contribution of different matching algorithms both to situations where the structure is decisive and where attributes are key. However, in localization, where the attributes are given by the application and can not be modified, we focus on the analysis of structural variability. We will find that the contribution of structure to image categorization in localization contexts is tightly bounded by the structural variability of the environment.

The rest of the paper is organized as follows. In Section 2 we describe two different similarity measures between attributed graphs extracted from the images, and proper graph-matching algorithms (graph-transformation matching and attributed Softassign) for computing them. These measures are the core of the graph clustering cost function described in Section 3, where we detail how the combination of an EM clustering algorithm and a supergraph-based prototype strategy improves structural learning. We choose an EM algorithm for the *chicken and egg* problem of determining prototypes after the clusters and correct clusters after the prototypes. Then, Section 4 is the core of the paper and it is there where we compare the contribution of matching algorithms to two different contexts of image categorization, and also where we discuss the added value of structural categorization. Finally, in Section 5 we present our conclusions and future works.

2. Similarity measures and graph matching

2.1. Size of consensus graph and GTM

Let $\mathbf{G}_i$ and $\mathbf{G}_j$, where $|V_i| = |V_j| = M$, be two node-attributed *K-NN* graphs after computing one-to-one putative matchings using β_i and β_j, respectively. Then, given their adjacency matrices A and B we define the *consensus graph* as the *K-NN* graph $\mathbf{G}_c = (V_c, E_c)$ where $v \in V_c \iff \mathcal{M}(k,l) = 1$ through the one-to-one matching $\mathcal{M}$, being $k \in V_i$, $l \in V_j$ and satisfying $A_{ka} \equiv B_{lb}\ \forall a \in KNNeighbor(k),\ \forall b \in KNNeighbor(l)$. Therefore, the matching induces an isomorphism between the *K-NN* graphs derived for selecting a given set of vertices in V_i and the same for V_j. Then, the latter matching also yields a good similarity measure $F_{ij} = |V_c|$ between the two input graphs.

Considering the two sets of M characteristic points $\widehat{\mathscr{P}}_i \ni \mathbf{s}_k$ and $\widehat{\mathscr{P}}_j \ni \mathbf{p}_l$, where $\mathbf{s}_k$ matches $\mathbf{s}_l$, we build their associated *median K-NN graphs* as follows. Graph $\mathbf{G}_i = (V_i, E_i, \beta_i)$ is given by vertices V_i associated to the positions of the M points. A non-directed edge $\langle k,a\rangle$ exists in E_i when $\mathbf{s}_a$ is one of the $K = 4$ closest neighbors of $\mathbf{s}_k$ and also $\|\mathbf{s}_k - \mathbf{s}_a\| \leqslant \eta$, being $\eta = med_{\langle r,t\rangle \in V_i \times V_i} \|\mathbf{s}_r - \mathbf{s}_t\|$ the median of all distances between pairs of vertices. The median filters structural deformations due to outlying points. If there are not K vertices that support the structure of $\mathbf{s}_k$ then this vertex is disconnected completely. The graph $\mathbf{G}_i$, which is not necessarily connected, has the $M \times M$ adjacency matrix A_{ka} where $A_{ka} = 1$ when $\langle k,a\rangle \in E_i$ and $A_{ka} = 0$ otherwise. Similarly, the graph $\mathbf{G}_j = (V_j, E_j, \beta_j)$ for points $\mathbf{p}_l$ has adjacency matrix B_{lb}, also of dimensions $M \times M$ because of the one-to-one initial matching $\mathcal{M}$. Each node is attributed with a feature vector β_j (SIFT descriptors. [24]).

Graph transformational matching [6,7] relies on the hypothesis that outlying matchings in $\mathcal{M}$ may be iteratively removed: (i) select an outlying matching; (ii) remove matched features corresponding to the outlying matching, as well as this matching itself; (iii) recompute both *median K-NN graphs*. Structural disparity is approximated by computing the residual adjacency matrix $R = |A - B|$ and selecting $j^{out} = \arg\max_{j=1\ldots M} \sum_{i=1}^{M} R_{ij}$, that is, the one which maximizes the number of different edges in both graphs. The selected structural outliers are the features forming the pair $(\mathbf{s}_k, \mathbf{p}_{j^{out}})$. Thus, we remove matching (k, j^{out}) from $\mathcal{M}$, $\mathbf{s}_k$ from $\widehat{\mathscr{P}}_i$, and $\mathbf{p}_{j^{out}}$ from $\widehat{\mathscr{P}}_j$. Then, after decrementing M, a new iteration begins, and the median K-NN graphs are computed from the surviving vertices. The algorithm stops when it reaches the null residual matrix, that is, when $R_{ij} = 0\ \forall\ ij$; it seeks a *consensus graph* (initial experimental evidence shows that the pruning with the residual adjacency matrix may be too aggressive). Considering that the bottleneck of the algorithm is the re-computation of the graphs, which takes $O(M^2 \log M)$ (the same as computing the median at the beginning of the algorithm) and also that the maximum number of iterations is M, the worst case complexity is $O(M^3 \log M)$. GTM is similar to bipartite graph matching because it is mainly cost driven although structure is somewhat decisive.

2.2. Normalized cost and attributed softassign

Alternatively, we define a symmetric and normalized dissimilarity measure between two graphs relying on the maximization of a cost function for graph matching:

$$F_{ij} = 1 - \frac{\max_{\mathcal{M}}[F(\mathbf{G}_i, \mathbf{G}_j; \mathcal{M})]}{\max[F_{ii}, F_{jj}]}, \tag{1}$$

$F_{ii} = F(\mathbf{G}_i, \mathbf{G}_i; I_{|V_i|})$, $F_{jj} = F(\mathbf{G}_j, \mathbf{G}_j; I_{|V_j|})$, being $I_{|V_i|}$ and $I_{|V_j|}$ the identity matrices defining self-matchings. This results in a normalized measure $F_{ij} = F_{ji} \in [0,1]$. That is, when the compared graphs are isomorphic, this value is 1, and as they become different this value is closer to 0.

A classical way of defining the cost function is to use the quadratic Gold and Rangarajan [15] one (adapted in this case to combine structural similarity with pairwise attribute similarity):

$$F(\mathbf{G}_i, \mathbf{G}_j; \mathcal{M}) = \sum_{k=1}^{|V_i|} \sum_{i=1}^{|V_j|} \sum_{b=1}^{|V_i|} \sum_{j=1}^{|V_j|} \mathcal{M}_{kl} \mathcal{M}_{ab} A_{ka} B_{lb} C_{kalb}, \tag{2}$$

where $\mathcal{M}$ is the matching matrix being evaluated, and C_{kalb} is a dissimilarity measure (e.g. decay of Euclidean distance) between the attributes of the nodes k and l, and a and b. The latter cost function is maximized by a procedure known as *graduated assignment*, which iteratively proposes a benefit matching matrix and