Pattern Recognition 74 (2018) 266–285

Contents lists available at ScienceDirect

Pattern Recognition

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

Graph-based bag-of-words for classification

Fernanda B. Silva[a,1], Rafael de O. Werneck[a,*], Siome Goldenstein[a], Salvatore Tabbone[b], Ricardo da S. Torres[a]

[a] RECOD Lab, Institute of Computing (IC), University of Campinas (Unicamp) Av. Albert Einstein, 1251, Campinas 13083-852, SP, Brazil
[b] Université de Lorraine-LORIA UMR 7503 BP 239, 54506 Vandoeuvre-lès-Nancy, France

ARTICLE INFO

Article history:
Received 6 September 2016
Revised 25 July 2017
Accepted 8 September 2017
Available online 12 September 2017

Keywords:
Bag models
Graph matching
Graph-based retrieval
Graph-based classification

ABSTRACT

This paper introduces the *Bag of Graphs (BoG)*, a Bag-of-Words model that encodes in graphs the local structures of a digital object. We present a formal definition, introducing concepts and rules that make this model flexible and adaptable for different applications. We define two BoG-based methods – *Bag of Singleton Graphs (BoSG)* and *Bag of Visual Graphs (BoVG)*, which create vector representations for graphs and images, respectively. We evaluate the Bag of Singleton Graphs (BoSG) for graph classification on four datasets of the IAM repository, obtaining significant results in accuracy and execution time. The method Bag of Visual Graphs (BoVG) is evaluated for image classification on Caltech and ALOI datasets, and for remote sensing image classification on images of Monte Santo and Campinas datasets. This framework opens possibilities for retrieval, classification, and clustering tasks on large datasets that use graph-based representations impractical before due to the complexity of inexact graph matching.

1. Introduction

Huge volumes of digital data have been created due to advances in acquiring, storing, sharing, and managing technologies. In this scenario, the appropriate use of data depends on the development of effective and efficient classification and retrieval tools, which in turn require the design of discriminant representation models of objects that enable us to identify/encode their similarities.

Bag-based representations have been extensively used to compute the similarity among digital objects by characterizing the frequency of occurrence of object features, *Bag of Words (BoW)* being one of the first successful models to create a vector representation of textual documents based on the frequency of word occurrences [1]. The adaptation of BoW for image context [2] is called *Bag of Visual Words (BoVW)*, or *Bag of Features*. This approach represents an image as a collection of visual words, where each visual word refers to a relevant visual pattern. The image descriptor is created based only on the number of occurrences of some particular visual appearances within the image.

This BoW model is a simple and efficient form of representation that enables a fast computation of object similarities. However, recent studies [3–5] have investigated the use of spatial information to improve these representations. Including local structures into the object description process can improve the bag representation, and lead to improvements in several tasks, which are dependent on the identification of semantic similarities. In fact, semantic meaning is a subjective concept, and it is not easily mapped to many digital objects. Sometimes we can use as evidence of similarity between a pair of objects the presence of *similar patterns* within them. These patterns may be defined in terms of relationships among object components, like spatial proximity. Thus, the use of a representation that describes an object through its local structures can lead to effective solutions for the recognition and categorization of digital objects. In this sense, graphs are a flexible tool for modeling relationships, and they are particularly useful for representing local structures within a digital object. Additionally, the invariance of graphs to several geometric transformations allows the creation of robust representations.

The hypothesis we explore in this paper is that the combination of graphs with the BoW model can create a discriminant and efficient representation based on local structures of an object, leading to fast and accurate results in classification tasks. The rationale is that the two representations are complementary and can help each other overcome their individual deficiencies. Graphs can encode local structures into a BoW-based descriptor, which can improve bag

* Corresponding author.
E-mail addresses: fernanda.silva@students.ic.unicamp.br (F.B. Silva), rafael.werneck@ic.unicamp.br (R.O. Werneck), siome@ic.unicamp.br (S. Goldenstein), tabbone@loria.fr (S. Tabbone), rtorres@ic.unicamp.br (R.S. Torres).
[1] Present address: Microsoft ATL, Rio de Janeiro, Brazil

http://dx.doi.org/10.1016/j.patcog.2017.09.018
0031-3203/

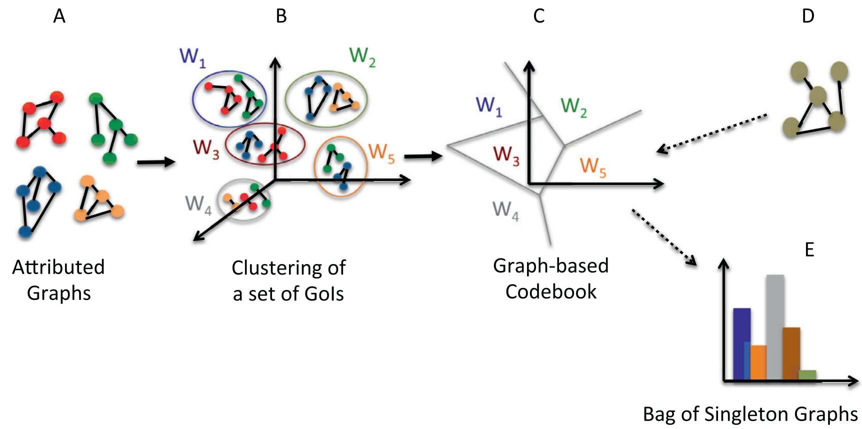

Fig. 1. Overview of the Bag of Singleton Graphs. We describe a set of graphs (A) in terms of vertex signatures, cluster them (B) to build the codebook (C), and count codeword occurrences within an input graph (D) to create the bag representation (E).

representations. At the same time, the use of BoW-based representations reduces the amount of time required by graph-based methods to compute the similarity between objects.

This paper introduces a novel object descriptor that combines bag and graph representations. We propose the *BoG*, a generic approach that creates a vector representation based on local structures defined by graph elements. This theoretical framework may be adapted to different contexts, and in this paper, we further describe two concrete realizations of the generic framework.

The first approach, called *BoSG*, generates a bag representation for objects that were previously modeled as graphs with attributes associated with their vertices and edges (illustrated in Fig. 1). The second approach, denominated *BoVG*, creates BoW-based descriptors using graphs to model the spatial relationships between the visual words found within an image. We also discuss the use of *BoVG* in the creation of a graph-based visual representation for remote sensing images that models the spatial relationships among their labeled regions. All case studies obtain accuracy rates comparable to other methods of the literature when evaluated on standard datasets [6–8].

This paper extends the works presented in [9,10]. Those papers present the use of the Bag-of-Graphs models concerning graph and image object classification problems. None of them, however, provides a comprehensive formal description of the model, which may guide researchers and developers in the creation of novel realizations and extensions. Another novelty of this journal paper refers to the introduction of a novel realization of the proposed model in the context of remote sensing image (RSI) representation and classification tasks. Finally, the experimental protocol was extended in order to include experiments with the ALOI dataset (in the case of image object classification) and model validation in RSI classification tasks. In summary, the main contribution of this work is the formal description of a generic graph model for digital object representation, with substantial practical demonstration through the instantiation, implementation, and validation of the theory in three real and distinct problems.

2. Related work

Graph is an abstract structure [11] that can be easily adapted, allowing its application in domains that range from biology to engineering [12,13]. Graphs can capture the relationships of an object's internal parts while being invariant to some transformations [14].

2.1. Graph representation

Some examples in image representation are the *graph of interest points* [15,16], *the graph of adjacent regions* [17,18], *the skeleton graph* [19–22], *the graph of primitives* [23,24], and *the graph of face fiducial points* [25].

In object recognition, *Spatial Relational Graphs (SRGs)* [24] describe symbols based on topological relationships of graphic primitives, while *Attributed Relational Graph (ARG)* can capture both topological and directional spatial relationships [23]. *Spatial Orientation Graph (SOG)* [26,27] describe the spatial positioning of objects within an image while *Skeleton Graphs* [28] and *Complete Graphs* [29] model the geometry of parts.

2.2. Graph matching

For most pattern recognition, indexing, and retrieval tasks, it is essential to compare similarities between data elements. So, when using graphs to represent objects, those tasks require the computation of the similarity between pairs of graphs, a complex problem usually addressed by graph-matching approaches, either exact or inexact.

Exact graph-matching algorithms determine if two graphs are isomorphic, a bijection between the elements of a pair of graphs. The complexity of exact graph matching has not yet been proven [30], but there are polynomial algorithms for solving the isomorphism problem of special types of graphs [12].

Inexact graph-matching algorithms provide a distance value that indicates graph dissimilarity. Different from the exact graph matching, the complexity of this problem has been proved to be NP-complete [30].

Graph-edit distance [31] is one of the most popular methods to perform inexact graph matching. Inspired by the traditional edit distance function that computes the similarity between two strings, the distance between a pair of graphs is defined as the minimum cost for converting one graph into another. This method is accurate, but it has an exponential time complexity [12]. Different Edit-Distance approaches propose sub-optimal edit cost with reduced computation time [32–34].

The use of traditional graph-matching methods to search and classify graphs on large datasets has severe limitations due to their high computational cost.

2.3. Graph embedding

The Vector Space Model (VSM) [35] is a well-known technique, commonly used in the context of text retrieval, that represents a