ARTICLE IN PRESS

J. Vis. Commun. Image R. xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Contextual aerial image categorization using codebook ☆

Zhijun Meng [a], Yan Wang [b], Xinyu Wu [c,*], Yating Yin [b], Teng Li [b]

[a] Beihang University, Beijing, China
[b] Anhui University, Hefei 230601, China
[c] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

ARTICLE INFO

Article history:
Received 21 September 2016
Revised 28 December 2016
Accepted 9 February 2017
Available online xxxx

Keywords:
Aerial image
Contextual modeling
Efficient algorithm
Codebook

ABSTRACT

Effective categorization of the millions of aerial images from unmanned planes is a useful technique with several important applications. Previous methods on this task usually encountered such problems: (1) it is hard to represent the aerial images' topologies efficiently, which are the key feature to distinguish the arial images rather than conventional appearance, and (2) the computational load is usually too high to build a realtime image categorization system. Addressing these problems, this paper proposes an efficient and effective aerial image categorization method based on a contextual topological codebook. The codebook of aerial images is learned with a multitask learning framework. The topology of each aerial image is represented with the region adjacency graph (RAG). Furthermore, a codebook containing topologies is learned by jointly modeling the contextual information, based on the extracted discriminative graphlets. These graphlets are integrated into a Bag-of-Words (BoW) representation for predicting aerial image categories. Contextual relation among local patches are taken into account in categorization to yield high categorization performance. Experimental results show that our approach is both effective and efficient.

1. Introduction

With the development of unmanned airplanes, aerial image categorization is becoming a more and more important topic. It has wide applicable scenarios such as video surveillance and abnormal detection. Furthermore, as captured from a totally different viewpoint, aerial image can be considered as a different modality from our daily-life captured photos, and its information can be complimentary to normal visual recognition system. For example, a scene annotation system can recognize the place as a downtown area more effectively with aerial image categorization.

In the past decade, Bag-of-Words (BoW) has become a popular technique in vision tasks such as visual search [1,2], object categorization [3,4], and action recognition [5,6]. It is originally from the text categorization area, and it shows to be effective and flexible. The basic computational flowchart is as the follows: extract local features from training images, build a visual vocabulary by clustering the local features to visual words, and then encode each local feature by mapping to a visual word, and represent an image as a vector containing the count of each visual word in that image [7].

However, to apply the existing codebook based BoW algorithm to aerial image categorization directly is problematic, due to such following reasons:

(1) Conventional codebook based methods usually are based on local appearance features, such as the local patches based scale invariant feature transform (SIFT), which cannot well discriminate different categories of aerial images. Fig. 1 shows some examples of aerial images, from which we can see it is hard to discriminate categories like "park", "industrial", "living area" based on their local appearance. The key feature that discriminates the aerial image categories is the various spatial layouts. The local areas of some aerial images could have quite similar appearances while their topologies are apparently different.
(2) The computational load is still too high to be applicable. It is well known that computational efficiency is critical for many applications. For example, it is necessary to process several aerial images in realtime when searching abnormal areas from aerial images. The conventionally adopted dense local patches sampling strategy and finely designed feature descriptors require high computational load [8]. Moreover, effective classifiers such as random forest bring more complexity [9].

☆ This paper has been recommended for acceptance by Anan Liu.
* Corresponding author.
E-mail address: xy.wu@siat.ac.cn (X. Wu).

http://dx.doi.org/10.1016/j.jvcir.2017.02.003
1047-3203/© 2017 Published by Elsevier Inc.

Fig. 1. Some exemplar aerial images.

(3) It can be seen that in aerial images contextual information for related local regions can improve the generalization performance. In [10], the contextual relations among local visual patches have been represented and incorporated into the codebook based framework, and it improves visual categorization performance greatly. In aerial images categorization, it is also important to discover the local patch contexts in the codebook.

Addressing the above mentioned issues, in this paper, aiming at categorizing aerial images effectively and efficiently, we propose a contextual topology-aware model based on the codebook. The key technique includes how to represent the local topologies in aerial images and encode them into an efficient codebook. Moreover, codebook based contextual modeling is considered here for effective aerial images categorization.

Fig. 2 shows the flowchart of the proposed aerial image categorization algorithm. Firstly, the images are segmented into atomic local regions. To extract the topological features, i.e., spatial layout information, the segmented regions are utilized to construct a region adjacency graph (RAG), with each local region as a vertex and spatially neighboring vertices are linked with edges. Although aerial image categorization can be implemented by comparing the RAG of a test aerial image to those of the labeled training aerial images, it is obviously computationally intractable. Therefore, in the next step, for the consideration of computational load, the graphlets with discriminative topologies are converted to a codebook based representation for comparison. The multiple task learning (MTL) framework [11] is adopted here to learn a codebook. It can learn the topological codebook that is discriminative for multiple aerial image categories. The contexts among local graphlets are further modeled with the contextual BoW [10]. Finally, a classifier is applied to predict aerial image categories based on the contextual BoW representations.

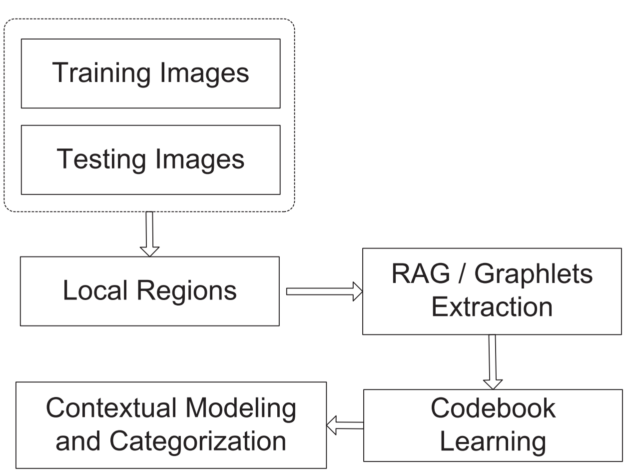

Fig. 2. Pipeline of the proposed codebook based aerial images categorization system.

The main contributions of this paper are: (1) we propose a codebook based representation based on the topological information that can classify aerial images efficiently and effectively; (2) we provide a contextual modeling framework that incorporates the local regions relation for better performance.

The rest of this paper is organized as follows. Section 2 gives the related works of contextual codebook based methods for aerial images categorization. Section 3 specified the proposed framework, including graphlets based feature extraction and codebook learning, contextual modeling and categorization methods. Section 4 describes the details of experiments, which thoroughly demonstrate the efficiency and effectiveness of the proposed method. Section 5 gives the conclusions.

2. Related work

In this part, the previous related work to this paper will be discussed in two categories according to the topics: codebook based categorization techniques and graph-based contextual image modeling.

In recent years, several works have been proposed on BoW based methods. Some works focused on the codebook learning step, aiming at better performance or higher efficiency. [12] gives a comparative study. Different clustering techniques, such as mean-shift [13] or hierarchical k-means [14], have been explored in visual vocabulary learning. But these are generative methods, discriminative visual words learning has also been studied to improve the classification performance. A compact and discriminative codebook learning was proposed by Winn et al. [15], which pair wise merges visual words based on the information bottleneck principle. The randomized forest was applied in codebook learning in Moosmann et al. [16]. A recent work [17] proposed to learn the codebook by minimizing information loss. The above mentioned works usually try to learn the visual words more efficiently, to better categorize images more accurately with high speed.

Another group of works tried to minimize the gap between low-level based visual words and the high-level semantic concepts. The middle level semantic topics extraction methods have been proposed based on BoW, based on which categories can be modeled [18,19]. The learning algorithms can be unsupervised and probabilistic latent semantic analysis (pLSA) or latent dirichlet allocation (LDA). Experiments also have been conducted on human action categorization [20] and video object discovery [21].

Simple geometric layout information has also been incorporated in later work [22]. Images are partitioned into increasingly fine grids and histograms are computed for patches found inside each grid cell, based on which the pyramid matching is adapted for classification, named Spatial Pyramid Matching (SPM). It also shows better categorization performance than above mentioned methods. Some other recent works try to incorporate the multimodal information in image categorization [23]. However, the spa-

Please cite this article in press as: Z. Meng et al., Contextual aerial image categorization using codebook, J. Vis. Commun. (2017), http://dx.doi.org/10.1016/j.jvcir.2017.02.003