J. Vis. Commun. Image R. 37 (2016) 53–62

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

An aerial image recognition framework using discrimination and redundancy quality measure

Yuxing Hu [a,*], Liqiang Nie [b]

[a] School of Aerospace, Tsinghua University, China
[b] School of Computing, National University of Singapore, Singapore

ARTICLE INFO

Article history:
Received 13 November 2014
Accepted 8 April 2015
Available online 21 April 2015

Keywords:
Aerial image
Categorization
Discriminative
Subgraph
Data mining
Image recognition
Framework
Quality measure

ABSTRACT

Aerial image categorization plays an indispensable role in remote sensing and artificial intelligence. In this paper, we propose a new aerial image categorization framework, focusing on organizing the local patches of each aerial image into multiple discriminative subgraphs. These meaningful subgraphs reflect both the geometric property and the color distribution of an aerial image. First, each aerial image is decomposed into a collection of regions in terms of their color intensities. Thereby region connected graph (RCG), which models the connection between the spatial neighboring regions, is constructed to encode the spatial context of an aerial image. Second, a novel subgraph mining technique is adopted to discover the frequent structures in the RCGs constructed from the training aerial images. Thereafter, a set of refined structures is selected among the frequent ones toward being highly discriminative and low redundant. Lastly, given a new aerial image, its sub-RCGs corresponding to the refined structures are extracted. They are further quantized into a discriminative vector for SVM classification. Thorough experimental results validate the effectiveness of the proposed method. In addition, the visualized mined subgraphs show that the discriminative topologies of each aerial image are discovered.

1. Introduction

Aerial image categorization is an important component for many applications in artificial intelligence and remote sensing [18,19,20,26], such as visual surveillance, navigation, and robot path planning. However, it is still a challenging task to deal with aerial image categorization successfully due to two reasons. On one hand, the aerial image components (*e.g.*, house roofs and grounds) as well as their spatial configurations are complex and inconstant, making it difficult to extract features sufficiently discriminative for aerial image representation. On the other hand, the efficiency of the existing aerial image categorization methods is far from practical due to the huge number of various components as well as their bilateral relationships. Therefore, a discriminative and concise aerial image representation has become increasingly imperative for a successful categorization system.

In the literature of designing discriminative image representations for visual recognition, many features have been proposed. They can be categorized into two groups: global features and local features. Global features, such as histograms, eigenspace [1], and skeletal shape [2], generalize the entire image with a single vector and are standard for statistic models like SVM. However, global features are sensitive to occlusion and clutter. Besides, these representations typically rely on a preliminary segmentation of objects in images. These two limitations result in unstable categorization performance. Different from global features, local features are developed to increase the discrimination, such as scale invariant feature transform (SIFT) [3]. Each local feature describes a localized image region and is calculated around the interest points. Thus, they are robust to partial occlusion and clutter. To take advantage of this property, local features [11–13] (*e.g.*, junction [4], gradient [5], contour, etc.) are widely used for aerial image parsing recently. However, when employing local features for image categorization, different images typically contain different numbers of local features. That is, it is difficult to integrate the local features within an image for the standard classifiers. In many cases, they are integrated into an orderless bag-of-features as global representation, thereby the similarity between images is determined by the orderless bag-of-features. It is worth emphasizing that as a non-structural representation, the bags-of-features representation ignores the geometric property of an image (*i.e.*, the spatial distribution of the local image patches), which prevents it from being highly discriminative. Fig. 1 depicts the significance of the geometric property. Given the zebra skin and the chessboard skin (Fig. 1(a1) and (b1)), their bag-of-features representations are

* Corresponding author.
E-mail address: thuyuxinghu@gmail.com (Y. Hu).

http://dx.doi.org/10.1016/j.jvcir.2015.04.004
1047-3203/

similar (Fig. 1(a3) and (b3)). That is to say, the bag-of-features representation is not sufficiently descriptive to distinguish the zebra and the chessboard, although the geometric properties of the two images are significantly different (Fig. 1(a2) and (b2)).

In order to encode image geometric proprieties into a categorization model, several image geometric features have been proposed. In [14], the spatial pyramid matching kernel is obtained by clustering the local features into a few geometric types. However, the spatial pyramid matching kernel is not flexible enough, since it highly depends on the human prior knowledge. RGB-domain spin image [15] describes the spatial context by exploring the chain structure of pixels in each RGB channel. However, the chain structure usually fails to describe the spatial context with complicated structures. Walk kernel [16] is proposed to capture the walk structures among image local features. However, the unavoidable totter phenomenon (*i.e.*, one vertex may occur several times in a walk) brings noise and hence limiting its discrimination. To obtain a better discrimination, parameters are provided to tune the length of the chain [15] or walk [16]. This operation leads to very redundant structures. Both the time consumption and the memory cost increase remarkably as the structure number goes up. Therefore, a concise image structure representation is desired for accurate aerial image categorization. Recently, many graph-based models are applied in intelligence systems and multimedia. They can be used as geometric image descriptors [38,39,27,28] to enhance image categorization. Besides, these methods can be used as image high-order potential descriptors of superpixels [29–32,6]. Further, graph-based descriptors can be used as a general image aesthetic descriptors to improve image aesthetics ranking, photo retargeting and cropping [7–10].

In this paper, we propose a novel aerial image categorization system, which enables the exploration of the geometric property embedded in local features. An aerial image is represented by a graph, since the graph is a natural and descriptive tool to express the complicated relationships among objects. By defining region connected graph (RCG), we decompose an aerial image into a set of discriminative subgraphs. To capture discriminative relationships among RCGs, a structure refinement strategy is carried out to select highly discriminative and low redundant structures. Based on the refined structures, we extract sub-RCGs accordingly and all the sub-RCGs from an aerial image form the discriminative spatial context. Finally, a quantization operation transforms the discriminative spatial context into a feature vector for categorization.

The major contributions of this paper are as follows: (1) region connected graph (RCG), a graph-based representation that describes the local patches and their topology for an areal image; (2) a structure refinement algorithm that selects highly discriminative and low redundant structures among the training RCGs; and (3) an efficient isomorphism subgraph extraction component that acquires the corresponding sub-RCGs.

2. Region connected graph (RCG)

An aerial image usually contains millions of pixels. If we treat each pixel as a local feature, highly computational complexity will make aerial image recognition intractable. Fortunately, an aerial image can be represented by a collection of clusters because pixels are usually highly correlated with their neighboring ones. Each cluster consists of neighboring pixels with consistent color intensities. Thus, given an aerial image, we can represent it by a set of regions instead of millions of pixels. The neighboring relationships between regions define the spatial context of an aerial image. Naturally, we can model this representation as a labeled graph.

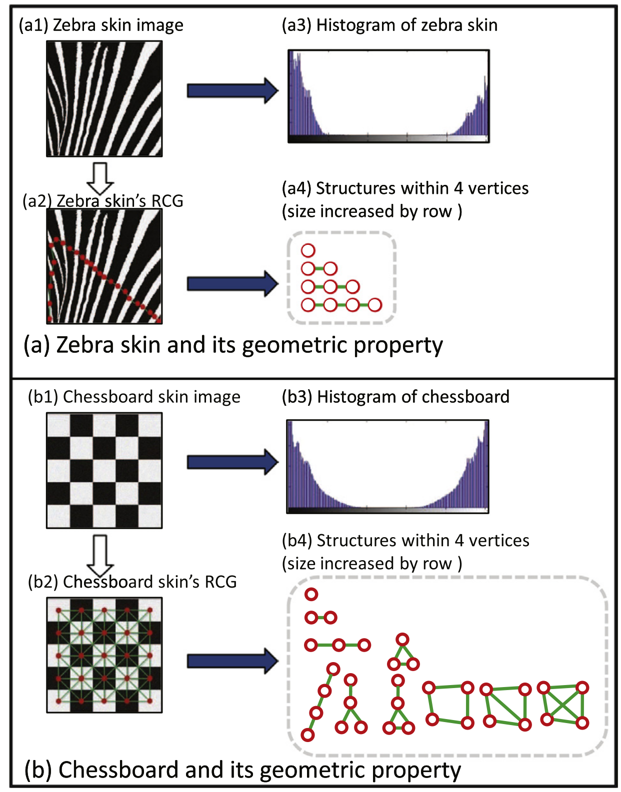

Fig. 1. The structures of zebra skin and chessboard (the red solid points are labeled vertices representing each region, the red hollow points are unlabeled vertices denoting each structure, and the green lines are edges). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The labels denote the local features of each region and each edge connects pairwise neighboring regions. In our work, we call this representation region connected graph (RCG).

To obtain the RCG from an aerial image, a segmentation algorithm (*i.e.*, fuzzy clustering [37] in our implementation) groups pixels into different clusters according to their color intensity. Note that the pixels in the same cluster are unnecessarily spatially neighboring. As shown in Fig. 2, we use different grayscale values to identify different clusters. Pixels in the face and the lower half of the Snoopy's body are grouped into the same cluster. However, it is more reasonable if they are categorized into different groups, since the face and the lower half of Snoopy are spatially isolated. To this end, a region growing algorithm [17] is employed to divide an image into regions iteratively. In each iteration, the region growing algorithm initializes the current region with a random pixel. It continues adding the spatially neighboring pixels in this region if the current pixel and the existing pixels come from

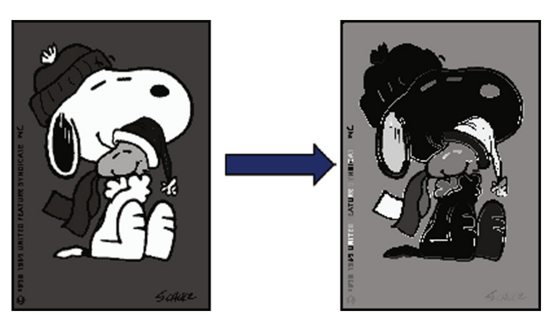

Fig. 2. From pixel clusters (left) to singly connected regions (right).