Pattern Recognition 60 (2016) 444–457

Contents lists available at ScienceDirect

Pattern Recognition

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

Spatial contextual superpixel model for natural roadside vegetation classification

Ligang Zhang [a,*], Brijesh Verma [a], David Stockwell [b]

[a] School of Engineering and Technology, Central Queensland University, Brisbane, Australia
[b] Department of Transport and Main Roads, Emerald, Queensland, Australia

ARTICLE INFO

Article history:
Received 15 July 2015
Received in revised form
6 May 2016
Accepted 12 May 2016
Available online 20 May 2016

Keywords:
Feature extraction
Image segmentation
Classification algorithms
Object recognition
Artificial neural networks

ABSTRACT

In this paper, we present a novel Spatial Contextual Superpixel Model (SCSM) for vegetation classification in natural roadside images. The SCSM accomplishes the goal by transforming the classification task from a pixel into a superpixel domain for more effective adoption of both local and global spatial contextual information between superpixels in an image. First, the image is segmented into a set of superpixels with strong homogeneous texture, from which Pixel Patch Selective (PPS) features are extracted to train class-specific binary classifiers for obtaining Contextual Superpixel Probability Maps (CSPMs) for all classes, coupled with spatial constraints. A set of superpixel candidates with the highest probabilities is then determined to represent global characteristics of a testing image. A superpixel merging strategy is further proposed to progressively merge superpixels with low probabilities into the most similar neighbors by performing a double-check on whether a superpixel and its neighour accept each other, as well as enhancing a global contextual constraint. We demonstrate high performance by the proposed model on two challenging natural roadside image datasets from the Department of Transport and Main Roads and on the Stanford background benchmark dataset.

1. Introduction

Reliably classifying the types of vegetation present in natural roadside video data is important for many real-world applications such as road risk assessment and identification, and vegetation growth monitoring and management. Approaches for vegetation classification in a natural image generally face the following challenges:

a) A high degree of unstructured, dynamic or even unpredictable configuration of vegetation (e.g. grasses can be various types with different color and texture).
b) Significant changes in environmental conditions (e.g. sunlight variations) depending on factors such as daytime, season, location and weather conditions.
c) A high dependence on image capturing settings such as camera configuration and resolution. The scene may be overexposed, underexposed or blurred.

For robust vegetation classification, one key question is how to extract robust and effective features that take into account both spatial and contextual information of objects at different scales [1,2]. Compared with pixel based grey or color features, spatial and contextual features have the advantage of capturing statistical information within a local neighborhood to represent the subtle appearance differences between objects, as well as the geometrical locations of objects to enforce spatial constraints on their distribution. There are several challenges for spatial and contextual feature extraction: a) the regional boundary problem, which means that extracting regularly sized patch features around pixels in regional boundaries unavoidably introduces noise into the extracted features. Most existing approaches perform patch based feature extraction (e.g. Gabor filters and SIFT) directly from the whole image (or region) without considering the possible impact of such noise. b) Contextual class-specific classification. Most existing work performs multiple-class classification for all classes, but class-specific classification coupled with contextual information can potentially produce boosted performance for specific applications. c) Region growing strategy. Existing algorithms iteratively merge neighboring pixels starting from the highest confident initial seed pixels, which suffer from the drawback of being highly dependent on the appropriate selection of those initial seeds, and thus they may have low reliability in natural conditions.

To address the above challenges, this paper proposes a Spatial Contextual Superpixel Model (SCSM) for natural roadside

* Corresponding author.
E-mail addresses: l.zhang@cqu.edu.au (L. Zhang), b.verma@cqu.edu.au (B. Verma), david.r.stockwell@tmr.qld.gov.au (D. Stockwell).

http://dx.doi.org/10.1016/j.patcog.2016.05.013
0031-3203/

vegetation recognition, which combines pixel based supervised class-specific classifiers and superpixel based unsupervised region merging for robust classification. To be specific, Pixel Patch Selective (PPS) features are designed to overcome the regional boundary problem, contextual class-specific classifiers are built to produce class probability maps for each class and an adaptive superpixel merging algorithm is proposed to alleviate the high dependence on initial superpixel seeds. The motivations of SCSM stem from the following points: a) superpixel algorithms [3], which group pixels into perceptually meaningful atomic regions, have advantages of coherent support regions for a single labelling on a naturally adaptive domain rather than on a fixed window, more consistent feature extraction capturing contextual neighboring information by pooling over feature responses from multiple pixels, and significant reduction of complexity and computation. b) Pixel intensity and color are able to reflect precise appearance characteristics of vegetation, which may be critical for recognizing vegetation with unpredictable characteristics in real-world conditions. c) Training and testing data contain critical and complementary information for vegetation classification. Features generated from training data represent general characteristics of all classes, while those from testing data reflect "local" characteristics of a testing image, such as lighting conditions and vegetation types (an example shown in Fig. 1). Thus methods combining them are expected to be able to lead to more robust classification automatically adaptive to the nature of the testing image.

The rest of the paper is organized as follows. Section 2 presents an overview of related work on vegetation classification and major contributions of this paper. Section 3 introduces the proposed framework. The details of the SCSM are described in Section 4. The experiments and result analysis are presented in Section 5. A discussion is presented in Section 6. Finally, the paper is concluded in Section 7.

2. Related work and contributions

This section reviews prior work on vegetation classification on roadside data and superpixel based algorithms. We also point out our contributions.

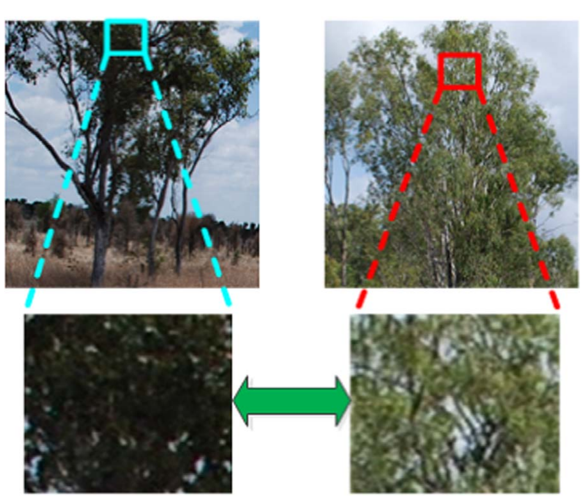

Fig. 1. Illustration of the importance of considering local characteristics of an object (i.e. tree) in a testing image. Because of different camera viewpoints to the sunlight direction, there are substantially contrasting differences in pixel intensities between the two tree regions. The tree types are also different. Thus a classifier trained using features from the left region is unlikely to be generalized well to the right region.

2.1. Related work

Vegetation classification on roadside data is a relatively less researched field, compared to a large amount of research using satellite and aerial data in the fields of remote sensing [4], agriculture, ecosystem, etc. It is closely related to image parsing [2] and scene labeling [5], which aim to find the region where a specific object presents and the category of the object belongs to. This paper reviews only closely related representative work, and for comprehensive surveys, readers are referred to [6,7].

Existing approaches can be broadly grouped into two categories – visible or invisible approaches. Visible approaches analyze the characteristics of vegetation in the visual spectrum and seek to utilize the color, shape, texture, geometry, and structure characteristics to distinguish vegetation from other objects, such as soil, tree, car, road, etc. A major benefit of visible features is that they retain high consistency with human visual perception, and they roughly fall into color and texture categories. Color is one of most popular features in existing research [8], which mainly focuses on investigating the suitability of different types of color spaces (e.g. Lab [9], YUV [10], HSV [11], and RGB [12]), while texture features are often represented by performing wavelet filters (e.g. Gabor filters [13] and Continuous Wavelet Transform [11]) or extracting pixel intensity distribution information in a neighborhood (e.g. pixel intensity differences [9,10] and variations [14,15]). The motion between video frames captured by optical flow has also recently been used to assist vegetation detection [11,16]. However, most of these approaches focus heavily on binary classification of vegetation versus non-vegetation. The way of extracting texture features does not take into account the boundary pixel problem, and the evaluation is largely limited to standard data without fully taking into account various challenges in natural scenes.

Invisible approaches focus on the use of the spectral properties of chlorophyll-rich vegetation and their reflectance characteristics in the invisible spectrum (e.g. infrared) to differentiate vegetation from other objects. For instance, a simple pixel-by-pixel comparison between red and Near Infrared Ray (NIR) reflectance potentially provides a powerful and robust way to detect photosynthetic vegetation [17]. The NIR has been modified to the Normalized Difference Vegetation Index (NDVI) [17], the modification of NDVI (MNDVI) [18], and combination of NDVI and MNDVI [16] to handle illumination effects such as shadow, shining, under- and over-exposure. The visible and invisible features have also been combined for improved results [13,14,19]. Compared to visible approaches, invisible approaches put high emphasis on Vegetation Indices (VIs) that are designed based on the reflectance characteristic of vegetation. However, the reliability of VIs for characterizing a vast amount of natural vegetation under challenging realistic conditions is still a question. In addition, invisible approaches often require the availability of specialized data capturing equipment such as ladar and laser, which limit their direct adoption in a wide range of applications.

There are also approaches that focus on general object classification, but they can be potentially used for vegetation classification such as texton based approaches [19,20], constrained parametric min-cut [21], hierarchical image model [2], hierarchical reconfigurable template [22] and others [23]. Recently, superpixel approaches, which take superpixels as the basic unit for object categorization, have shown good performance in extracting contextual and spatial features. Superpixels give better spatial support for aggregating features that could belong to the same object, and utilizing feature statistics on a naturally adaptive domain rather than on a fixed window. A work relevant to our research is presented in [24], which explored the use of superpixel neighborhoods for object identification by aggregating histograms of a