ISPRS Journal of Photogrammetry and Remote Sensing 122 (2016) 145–166

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

MRF-based segmentation and unsupervised classification for building and road detection in peri-urban areas of high-resolution satellite images

Ilias Grinias [a,*], Costas Panagiotakis [b], Georgios Tziritas [c]

[a] Department of Civil Engineering, Surveying and Geomatics, Technological Educational Institute of Central Macedonia, 62124 Serres, Greece
[b] Department of Business Administration, Technological Educational Institute of Crete, 72100 Agios Nikolaos, Greece
[c] Department of Computer Science, University of Crete, 70013 Heraklion, Greece

ARTICLE INFO

Article history:
Received 25 April 2016
Received in revised form 25 October 2016
Accepted 25 October 2016

Keywords:
Building/road extraction
Satellite images
Image segmentation
Feature analysis
Random Forest
Unsupervised classification

ABSTRACT

We present in this article a new method on unsupervised semantic parsing and structure recognition in peri-urban areas using satellite images. The automatic "building" and "road" detection is based on regions extracted by an unsupervised segmentation method. We propose a novel segmentation algorithm based on a Markov random field model and we give an extensive data analysis for determining relevant features for the classification problem. The novelty of the segmentation algorithm lies on the class-driven vector data quantization and clustering and the estimation of the likelihoods given the resulting clusters. We have evaluated the reachability of a good classification rate using the *Random Forest* method. We found that, with a limited number of features, among them some new defined in this article, we can obtain good classification performance. Our main contribution lies again on the data analysis and the estimation of likelihoods. Finally, we propose a new method for completing the road network exploiting its connectivity, and the local and global properties of the road network.

1. Introduction

Automatic detection of buildings and roads in aerial/satellite images is of great importance in a wide range of areas, such as urban planning, urban area monitoring/detection, change detection, construction and update of GIS maps, transportation and telecommunication. Such man-made structures appear with high density and regular patterns in scenes of urban areas, while, by contrary, in rural regions is not unusual to find only few buildings spread out over large distances and accessible by a sparse network of, often not paved, roads. Peri-urban areas (Ravetz et al., 2013) are defined as the transition zones where urban and rural uses mix. As in case of suburban regions, buildings are in neighborhoods and are surrounded by yards in varying densities and directions, while their road network usually follows relatively regular patterns and is often sparser than that of suburban areas.

1.1. Object detection in Remote Sensing Imagery (RSI)

Although aerial images have traditionally been used to extract buildings/roads for mapping applications (Mayer, 1999; Ahmadi et al., 2010; Hu et al., 2007; Mnih and Hinton, 2010), the successive launching of high spatial resolution commercial satellites IKONOS, QuickBird, WorldView (1, 2 and 3) and Geoeye-1, has led to high-resolution, cost-effective satellite imagery. One of the main difficulties of image processing tasks when moving from (either aerial or satellite) images with low (coarser than 10 m) and medium (of a few meters) resolutions to high (metric or sub-metric) resolution ones, is to be able to deal with the high complexity of the image content. This high complexity is mainly due to the fact that the elements or objects of interest are not any more only individual pixels or surfaces, but complex, structured groups of pixels (Inglada, 2007; Blaschke, 2010).

Objects under detection or localization may be man-made ones, such as vehicles, ships, buildings and roads, that have sharp boundaries and are not part of the background, as well as landscape objects, such as trees and land-use/land-cover parcels that often are not characterized by clear boundaries and hence, may be considered natural parts of the background environment. Even though

* Corresponding author.
E-mail addresses: elgrinias@gmail.com (I. Grinias), cpanag@staff.teicrete.gr (C. Panagiotakis), tziritas@csd.uoc.gr (G. Tziritas).

http://dx.doi.org/10.1016/j.isprsjprs.2016.10.010
0924-2716/

with the advances of remote sensing technology a greater range of man-made objects become separable from their background, the explosion in the availability of high-resolution Remote Sensing Imagery (RSI) underscores the need for automated satellite image interpretation methods. Such imagery has greatly increased the number of possible applications, but at the cost of an increase in the amount of required manual processing. Recent applications of large-scale machine learning to such high-resolution imagery (Inglada, 2007; Kluckner and Bischof, 2009; Paisitkriangkrai et al., 2015; Volpi and Ferrari, 2015; Vakalopoulou et al., 2015) have produced object detectors characterized by high levels of accuracy, reinforcing the belief that automated aerial/satellite image interpretation systems are within reach.

In RSI applications, aerial/satellite image interpretation is usually formulated as a pixel labeling task. Given an image, the goal is to produce either a complete semantic segmentation of the image into classes of objects such as "building", "road", "tree", "grass", and "water" or a binary classification of the image for a single object class. A very recent and complete review of object detection and localization methods in RSI, is found in Cheng and Han (2016). According to this review, the very large number of object detection methods can generally be divided into four, not necessarily independent, main categories: template matching-based methods, knowledge-based methods, Object Based Image Analysis (OBIA)-based methods, and machine learning-based methods. Among them, OBIA-based or Geographic OBIA (GEOBIA)-based methods (Blaschke, 2010; Blaschke et al., 2014) have become a very promising alternative for detecting objects in high-resolution (sub-meter) RSI. Those methods consist of two main parts, namely, image segmentation in "homogeneous" pixel regions or (hopefully) "meaningful" objects of interest, followed by feature classification of the resulting objects based on various extracted features of objects such as spectral information, texture, shape, size, geometry and semantic features (Blaschke et al., 2014).

1.2. Previous works on building and road network extraction

Several methods of the four categories have been proposed in literature for extracting the man-made objects and in particular road network and buildings, from satellite imagery of low, medium or high spatial resolution, using spectral or/and shape and structural (topology) properties, as it is described in detail in the following paragraphs. Several methods rely on supervised, ground truth based classification (Paisitkriangkrai et al., 2015; Vakalopoulou et al., 2015) and in some of them previously stored information, such as road centerlines in vector data format, is also assumed and used (Yuan and Cheriyadat, 2013).

Road and building detection are applied either on their own or simultaneously (Ünsalan and Boyer, 2005), since, depending on the methodology followed, the joint detection approach may improve the detection of both. An interesting, template-matching method has been proposed by Karantzalos and Argialas (2009), capable of detecting either road network or buildings by tuning the parameters of a level set (Sethian, 1999) algorithm.

Aytekin et al. (2012) detect buildings using spectral properties in conjunction with spatial properties, both of which provide complementary information to each other. Natural and man-made regions are classified and segmented using Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974; Myneni et al., 1995). Shadow regions are detected, and the rest of the image, consisting of man-made areas only, is partitioned by mean shift segmentation (Cheng, 1995; Comaniciu and Meer, 2002). Resulting segments, whose shape is irrelevant to that of buildings are eliminated using morphological operations. Karantzalos and Paragios (2009) introduced competing shape priors, and building extraction is addressed through a segmentation approach that involves the use of a data-driven term constrained from the prior models.

A method based on local feature point extraction using Gabor filters (Jain et al., 1997) is described in Sirmaçek and Ünsalan (2010). Local feature points vote for the candidate urban areas and final urban area is detected using an optimal decision-making approach on the vote distribution. In Sirmaçek and Ünsalan (2011) building detection is achieved using probability density functions of four, locally extracted, feature values. Finally, Benedek et al. (2012) introduced a global, probabilistic optimization process to find the optimal configuration of buildings, considering the observed data, prior knowledge, and interactions between the neighboring building parts. Since the method integrates building extraction with change detection in aerial and satellite imagery, the authors, apart from the results for change detection, provide quantitative performance results for building detection in a benchmark dataset that they created and consists mainly by the images of their freely available SZTAKI-INRIA building detection dataset.[1]

Recently, novel semantic labeling/segmentation techniques of high accuracy have been proposed, using Convolutional Neural Network (CNN) features (Jin and Davis, 2007; Wang et al., 2015) and Conditional Random Fields (CRFs) (Lafferty et al., 2001; Kumar and Hebert, 2003) to smooth region labeling, while respecting the edges of the image (Paisitkriangkrai et al., 2015). Volpi and Ferrari (2015) model the segmentation problem by a CRF as well, employing Structured Support Vector Machines (SSVM) (Tsochantaridis et al., 2005; Finley and Joachims, 2008) to learn both the weights of a set of visual descriptors and local class interactions. In Vakalopoulou et al. (2015) an automated building detection framework is proposed based on deep convolutional networks. The core of the developed method is based on a supervised classification procedure employing a very large training dataset. Using a Markov Random Field (MRF) model (Li, 2009) the classification result is improved. Experimental results are given on the data set used also in our work. Their quantitative validation indicates that this approach is quite promising.

A comprehensive review of automatic road network extraction techniques for GIS update is found in Mena (2003). In Laptev et al. (2000) roads are automatically extracted using the multi-scale detection of roads in combination with geometry-constrained edge extraction using ribbon snakes (Mille et al., 2008). Shadows and partially occluded areas are detected, as the bridges between partially (dis-)connected road segments and the road network is constructed after extracting crossings with varying shape and topology. In Huang and Zhang (2009), spectral and structural features are extracted in a number of scales and classified using Support Vector Machines (Vapnik, 1995). A majority voting approach is then used to integrate the multi-scale road information at the decision level in order to extract road centerlines and roads map.

In Das et al. (2011) a multistage framework for road network extraction is proposed by fusing region and boundary information to segment the image and then applying morphological operations to reject false positives. In Ünsalan and Sirmaçek (2012) probabilistic and graph theoretic methods are used to extract centerline and shape of roads.

Valero et al. (2010) build a granulometry chain using Path Openings and Path Closings (Talbot and Appleton, 2007) to construct Morphological Profiles. For each pixel, the Morphological Profile constitutes the feature vector on which the road extraction is based. In Hu et al. (2007), local homogeneous regions are enclosed by polygons, called footprints of pixels upon which road detection is achieved using tree expansion and pruning techniques.

[1] http://web.eee.sztaki.hu/remotesensing/building_benchmark.html.