Expert Systems With Applications 58 (2016) 130–142

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Building detection from orthophotos using a machine learning approach: An empirical study on image segmentation and descriptors

Fadi Dornaika[a,b,*], Abdelmalik Moujahid[a], Youssef El Merabet[c], Yassine Ruichek[d]

[a] University of the Basque Country UPV/EHU, Manuel Lardizabal 1, San Sebastian 20018, Spain
[b] IKERBASQUE, Basque Foundation for Science, Maria Diaz de Haro, 3, Bilbao 48013, Spain
[c] Faculté des Sciences, Université Ibn Tofail, Campus Universitaire, BP 133, Kénitra 14000, Morocco
[d] IRTES-SET, University of Technology of Belfort-Montbeliard, 90010 Belfort, France

ARTICLE INFO

Article history:
Received 22 May 2015
Revised 10 March 2016
Accepted 11 March 2016
Available online 31 March 2016

Keywords:
Automatic building detection and delineation
Orthophotos
Image segmentation
Image descriptors
Supervised learning
Classifier

ABSTRACT

Building detection from aerial images has many applications in fields like urban planning, real-estate management, and disaster relief. In the last two decades, a large variety of methods on automatic building detection have been proposed in the remote sensing literature. Many of these approaches make use of local features to classify each pixel or segment to an object label, therefore involving an extra step to fuse pixelwise decisions. This paper presents a generic framework that exploits recent advances in image segmentation and region descriptors extraction for the automatic and accurate detection of buildings on aerial orthophotos. The proposed solution is supervised in the sense that appearances of buildings are learnt from examples. For the first time in the context of building detection, we use the matrix covariance descriptor, which proves to be very informative and compact. Moreover, we introduce a principled evaluation that allows selecting the best pair segmentation algorithm-region descriptor for the task of building detection. Finally, we provide a performance evaluation at pixel level using different classifiers. This evaluation is conducted over 200 buildings using different segmentation algorithms and descriptors. The performance analysis quantifies the quality of both the image segmentation and the descriptor used. The proposed approach presents several advantages in terms of scalability, suitability and simplicity with respect to the existing methods. Furthermore, the proposed scheme (detection chain and evaluation) can be deployed for detecting multiple object categories that are present in images and can be used by intelligent systems requiring scene perception and parsing such as intelligent unmanned aerial vehicle navigation and automatic 3D city modeling.

1. Introduction

1.1. Motivation

Nowadays, automatic object recognition is a topic of growing interest for the machine vision community. In particular, the automatic building detection from monocular satellite and aerial images has been an important tool for many applications such as creation and update of maps and the Geographical Information Systems database, land use analysis, change detection and urban monitoring applications (Quang, Thuy, Sang, & Binh, 2015; Sirmacek & Unsalan, 2009; Sun et al., 2016; Unsalan & Boyer, 2005). Due to the rapidly growing urbanization, detecting buildings from images is a hot topic and an active field of research. Recently, vision and photogrammetry tools have been increasingly used in the processing of Geographical Information Systems, cultural heritage modeling, risk management, and monitoring of urban regions. More specifically, extracting objects such as roads and buildings has gain significant attention over the last decade. Aerial data are very useful for the coverage of large areas such as cities and several aerial-based approaches have been proposed for the extraction of buildings.

More precisely, the data employed as input to these approaches are either optical aerial images and derived Digital Surface Models (e.g., Tournaire, Brédif, Boldo, & Durupt, 2010) or aerial LiDAR 3D point clouds (e.g., Wang, Lodha, & Helmbold, 2006). It is well-known that segmenting buildings in aerial images is a challenging task. This problem is generally considered when we talk about high-level image processing in order to produce numerical or symbolic information. In this context, many techniques have

* Corresponding author at: University of the Basque Country UPV/EHU, Manuel Lardizabal 1, 20018 San Sebastian, Spain, Tel.: +34943018034.

E-mail addresses: fadi.dornaika@ehu.es, fdornaika@gmail.com (F. Dornaika), jibmomoa@gmail.com (A. Moujahid), elmerabet113@gmail.com (Y. El Merabet), yassine.ruichek@utbm.fr (Y. Ruichek).

http://dx.doi.org/10.1016/j.eswa.2016.03.024
0957-4174/

been proposed in the literature. Among the techniques most frequently used, one can cite semi-automatic methods that need user interaction in order to extract desired targets or objects of interest from images. Generally, this category of methods has been introduced to overcome the problems associated with the full automatic segmentation which is usually not perfect. It consists in dividing an image into two classes: "object" and "background". The interactivity consists in imposing some constraints to the segmentation that stipulate that some pixels (seeds) should belong to the object and some pixels should belong to the background. Rother, Kolmogorov, and Blake (2004) presented an iterative algorithm called GrabCut by simplifying user interaction. Their method combines image segmentation using graph cut and Gaussian mixture models (with the Orchard–Bouman clustering algorithm) of foreground and background structures in color space. A useful segmentation platform has recently been introduced by McGuinness and O'Connor (2010). The authors compared segmentation methods such as seeded region growing (SRG) (Adams & Bischof, 1994), Iterative Graph Cuts (Boykov & Jolly, 2001), and simple interactive object extraction (SIOX) (Friedland, Jantz, & Rojas, 2005).

1.2. Contribution

From the point of view of machine learning paradigms, it is desirable to keep the user interaction at the training phase only and to fully automate the detection and recognition at the test phase. In this paper, we propose an image-based approach for object detection and classification namely, detecting roof building in orthophotos. We use image segmentation algorithms to get an over-segmented orthophoto (e.g., Arbelaez, Maire, Fowlkes, & Malik, 2011; Nock & Nielsen, 2004). The obtained regions are then described by holistic and hybrid descriptors for detection of roof building in orthophotos. First, an over-segmentation is applied on the orthophoto. This over-segmentation is applied on both the training and test images. Second, holistic descriptors including color and texture are fused in order to get the feature descriptor of a given region. Third, the segmented regions in a test image are then classified using machine learning tools. We investigate the good combination (segmentation, descriptors) that can lead to optimal detection results via a case study over a set of aerial images. The main contributions of the paper are as follows. Firstly, we apply the matrix covariance descriptor to the building detection problem. To the best of our knowledge, this recent descriptor was not used in the context of building detection. This descriptor has proved to be very informative and compact. Secondly, we introduce a principled evaluation that studies the performances of the two main modules used in the detection chain, namely the image over-segmentation algorithm and the descriptor extractor. This study can provide and select the best pair segmentation algorithm-region descriptor in the context of building detection. Thirdly, we provide a performance study on classifiers whose role is to decide if any arbitrary region is a building or not. We provide evaluation performances over 200 buildings using different segmentation algorithms and descriptors.

While the application of the covariance descriptor to the building detection problem can be considered as one novel aspect of this current work, we point out that the objective of our work is not to propose a novel processing algorithm. Rather we are interested in studying the performance of a machine learning approach and its processing pipeline that combines several modules: image segmentation, image descriptor extraction, and classification. Thus, the work studies the influence of different modules on the final performance of building detection in orthophotos. Based on this study, we can identify the configurations that should be adopted for the task at hand.

The rest of the paper is organized as follows. Section 2 presents some related, state-of-the-art work. Section 3.1 describes the proposed machine learning approach as well as its main differences with existing work. It also presents the studied image descriptors together with their implementation details. Section 4 presents the performance evaluation through the use of image segmentation and descriptors classification. It presents an extended performance study of several combinations of pairs segmentation algorithm-image descriptor as well as of several classifiers. Finally, Section 5 provides some discussions and Section 6 concludes the paper.

2. Related work

This section is split in two main subsections. The first subsection describes briefly four image segmentation algorithms. The second subsection provides with an overview of the state-of-the art in building detection.

2.1. General purpose image segmentation

Before we proceed to the analysis of the effect of different segmentation methods on building roofs detection, we briefly review four popular segmentation techniques of the literature: Roerdink and Meijster (2001), Statistical Region Merging (SRM) (Nock & Nielsen, 2004), mean shift-based segmentation (MS) (Comaniciu & Meer, 2002), and Superpixels (Ren & Malik, 2003). These segmentation methods are well known and often used for building segmentation purposes. Most of these methods have several control parameters. Some parameters specify the image size or the output format. Other parameters are essential for the segmentation process.

Watershed algorithm: The Watershed algorithm is widely encountered and various definitions can be found in the literature (Roerdink & Meijster, 2001), (Vincent & Soille, 1991). In order to obtain the segmented image, the watershed technique uses a gradient image as input image, calculated using Di-zenzo's operators (Di-Zenzo, 1986) applied on the initial image. This segmentation technique has two main advantages: (1) it produces regions that are closed and connected; (2) the boundaries coincide with the most significant edges in the image.

Statistical Region Merging (SRM): Introduced by Nielsen and Nock, SRM is a fast and robust image segmentation technique (Nock & Nielsen, 2004). The algorithm considers each pixel as a region and it merges connected regions when their intensities are sufficiently similar according to a statistical test. This algorithm presents the advantage of not requiring any quantization or color space transformations. The number of regions is controlled by only a simple parameter Q, which represents the statistical complexity of the image and controls the level of segmentation.

Mean shift algorithm (MS): Based on gradient estimation, MS is an iterative statistical procedure to mode detection and clustering (Comaniciu & Meer, 2002). The mean shift approach is composed of two main steps: (1) filtering of the initial image and (2) clustering of the filtered image. The size and number of the obtained regions are controlled by two bandwidth parameters: the spatial bandwidth h_s related to the two spatial features and the range bandwidth h_r related to the color coordinate part of the feature vector. It is noted that the contours and the small regions are preserved after filtering.

Superpixels algorithm: Introduced by Ren and Malik (2003), Superpixels method is based on the graph cut algorithm that operates on graphs whose nodes are pixel values and whose edges represent affinities between pixel pairs. The advantage of the superpixels concerns their autonomous adaptation to the image structure where the produced regions, called "superpixels", are