ISPRS Journal of Photogrammetry and Remote Sensing 105 (2015) 107–119

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach

Shihong Du *, Fangli Zhang, Xiuyuan Zhang

Institute of Remote Sensing and GIS, Peking University, Beijing 100871, China

ARTICLE INFO

Article history:
Received 26 November 2014
Received in revised form 16 March 2015
Accepted 17 March 2015
Available online 14 April 2015

Keywords:
Very high resolution (VHR) images
Urban buildings
Semantic classification
Random forest
Object-based image analysis (OBIA)

ABSTRACT

While most existing studies have focused on extracting geometric information on buildings, only a few have concentrated on semantic information. The lack of semantic information cannot satisfy many demands on resolving environmental and social issues. This study presents an approach to semantically classify buildings into much finer categories than those of existing studies by learning random forest (RF) classifier from a large number of imbalanced samples with high-dimensional features. First, a two-level segmentation mechanism combining GIS and VHR image produces single image objects at a large scale and intra-object components at a small scale. Second, a semi-supervised method chooses a large number of unbiased samples by considering the spatial proximity and intra-cluster similarity of buildings. Third, two important improvements in RF classifier are made: a voting-distribution ranked rule for reducing the influences of imbalanced samples on classification accuracy and a feature importance measurement for evaluating each feature's contribution to the recognition of each category. Fourth, the semantic classification of urban buildings is practically conducted in Beijing city, and the results demonstrate that the proposed approach is effective and accurate. The seven categories used in the study are finer than those in existing work and more helpful to studying many environmental and social problems.

1. Introduction

As main sites of urban activities and important components of cities, urban buildings are vital foundations of urban studies. Semantic classification of buildings intends to label buildings using a set of semantic categories cognized and conceptualized by people, such as low-story shantytowns, middle-story apartments, high-story apartments, administrative buildings, and commercial buildings. These categories strongly correlate with urban environment analyses (e.g. ecological and environmental evaluation), urban resource allocation (e.g. resource management, transportation planning, and disaster reduction) and urban social analyses (e.g. population estimation, and market research) (Wu et al., 2005). Existing work has focused on how to extract building contours or accurately distinguish buildings from non-buildings. However, geometric information alone cannot fulfill the demands on urban ecology, resources and social researches (Paul et al., 2001). Therefore, semantic classification of urban buildings is required.

Geometric analyses of buildings have been intended to extract geometric contours of buildings or distinguish buildings from other objects by using geometric or spectral features. In the middle-to-late 1980s, researchers started to extract urban buildings from aerial photos (Huertas and Nevatia, 1988). With the explosive increase in image data and continuous development of sensor techniques, techniques of extracting urban buildings have made great progresses. From the perspective of images used, buildings can be extracted from either low- and medium-resolution images or high-resolution images (Lin and Nevatia, 1998). Due to the limits of spatial resolution, only large areas of buildings or residential areas instead of individual buildings can be obtained from low- and medium-resolution images (Nevatia et al., 1997). On the other hand, VHR images can provide finer texture and more accurate locations of buildings. Thus, they are used more comprehensively to acquire buildings in high accuracy (Myint et al., 2011). From the perspective of extraction methods, existing work generally falls into edge-based geometric grouping or object-based classification. The former first extracts edges from images, and then uses geometric models of buildings as prior constraints to find edges belonging to the same buildings and group them into complete contours. These works have often used optical VHR data (Kim and Muller,

* Corresponding author. Tel.: +86 10 62750294; fax: +86 10 62751961.
E-mail address: dshgis@hotmail.com (S. Du).

http://dx.doi.org/10.1016/j.isprsjprs.2015.03.011
0924-2716/

1999; Sirmacek and Unsalan, 2009; Ok, 2013) or a combination of optical and LiDAR data (Sohn and Dowman, 2007; Awrangjeb et al., 2010, 2013). Unlike edge-based methods, object-based methods first segment VHR images into image objects, and then distinguish image objects of buildings from that of non-buildings using image features (Myint et al., 2011). However, in VHR images, a lot of detailed information emerges, and the heterogeneity of buildings becomes much larger. Consequently, it is difficult to find appropriate segmentation scales and image features to classify complete buildings with different shapes, sizes, and structures.

Semantic analyses of urban buildings have concentrated much on distinguishing different categories of buildings. These categories are cognized and conceptualized by people and described by natural languages. More importantly, they are strongly correlated to environmental and social variables and have special implications to these variables. There have been a few studies concentrating on recognizing the categories or neighborhoods of urban buildings. For identifying the categories of buildings, Lu et al. (2014) used spatial attributes calculated from LiDAR and other land-use features to classify buildings into three categories: single-family houses, multiple-family houses, and non-residential buildings. Belgiu et al. (2014) used airborne laser scanning data to group buildings into three categories: residential/small buildings, apartments/block buildings, and industrial/factory buildings. For the classification of neighborhoods, Graesser et al. (2012) defined urban neighborhoods as homogeneous zones and classified them as formal and informal areas, but they did not recognize subtypes, such as residential, commercial, and industrial structures. Other work in this field includes extracting unplanned settlements (Kuffer et al., 2014) and slums (Kohli et al., 2012) from VHR images.

In terms of analyses above, most existing studies have focused on extracting geometric information on buildings while only a few have concentrated on semantic analysis. In addition, some important issues still remain to be resolved. First, existing work on semantic analyses has distinguished too few categories to satisfy the many demands in environmental or social sciences (Graesser et al., 2012; Kohli et al., 2012; Belgiu et al., 2014; Kuffer et al., 2014; Lu et al., 2014). Second, there have been no appropriate segmentation scales and algorithms to produce single image objects for diverse buildings. This lack of computational methods leads to low classification accuracies as image features strongly depend on segmentation scales. Third, a small number of manually chosen samples and features may be practical for classifying a few categories of buildings. To distinguish between more building categories greatly varying in size, shape, structure, and spectrum, however, a large number of samples, high-dimension and heterogeneous features are required. In this situation, the samples are often imbalanced, and the features are often auto-correlated and have distinct importance for distinguishing different categories. Unfortunately, there is still a lack of related work to reduce the influences of imbalanced samples on classification and to evaluate feature importance to classifying each category.

Aimed to resolve the issues raised above, this study presents a two-level segmentation mechanism (i.e. a large-scale layer constrained by GIS data for producing single image objects and a small-scale layer providing intra-object component features) and a semi-supervised method to choose a large number of unbiased samples by considering the spatial proximity and intra-cluster similarity of buildings. Random forest (RF) classifier is used to semantically classify buildings, for it is capable of handling a large number of samples and high-dimension and heterogeneous features. Moreover, to improve classification accuracy and evaluate feature importance, two improvements in RF classifier are presented: a voting-distribution-ranked rule for reducing the influences of imbalanced samples and a feature importance measurement for each category based on Gini descent and path tracing strategy.

The first contribution of this study is the improvements of RF classifier in voting rule and feature importance evaluation. Although some researchers have used RF classifier to classify VHR images, the effective approaches to handling imbalanced samples and evaluating feature importance for each category are still unresolved. The improvements fill the gap from a methodological perspective. Another contribution is the semantic classification of urban buildings. The seven categories used in this study are finer than those used in existing studies and more appropriate for many environmental and social variables, such as population distribution (Wu et al., 2005; Lu et al., 2006) and small-scale heating networks (Geiß et al., 2011). Existing categories are ineffective at handling these variables. This situation will be worse in China because the inter-category differences in the capability of holding families are very large. Therefore, this study is motivated by both theoretical and practical demands.

2. Semantic category system of buildings

This section will first discuss the cognition and representation of urban buildings in the real physical world, the geoinformatic world, and the cognition world, and then analyze the transformations of real-world urban buildings to object features and semantic categories. Finally, it will construct a semantic category system of buildings.

2.1. Category system of urban buildings

Urban buildings made of various materials with assorted styles and appearances in the real physical world are the basis of cognizing semantic category by people and of sensing buildings by remote sensors (the middle section of Fig. 1). In the geoinformatic world (the right section of Fig. 1), buildings are abstracted into contours in GIS data and into image pixels or image objects in VHR images. Thus, they are described from the aspects of spectrums, shapes, and textures. In the cognition world, people cognize, understand, and communicate their ideas about buildings through appropriate semantic categories (the left section of Fig. 1). Therefore, building a semantic category system helps to transform the feature representations in the geoinformatic world to the concepts in the cognition world.

The goal of semantic classification is to build relationships between the concepts of buildings in the cognition world and the features of buildings in the geoinformatic world. Therefore, the semantic category system can be built by discriminating the appearances and functions of urban buildings, including low-story (LS) shantytowns, medium-story (MS) apartments, high-rising (HR) apartments, administrative (AD) buildings, commercial (CM) buildings, industrial (ID) buildings, and auxiliary (AU) buildings (Table 1).

2.2. Inter-category variations of buildings

Fig. 2 illustrates seven typical images for each category of buildings. It is clear that these categories greatly vary in the following aspects. First, the buildings in those categories have different sizes. Most buildings are single objects, while some (e.g. LS shantytowns) refer to the extents of spatially dense buildings and are much larger than other buildings. Even buildings in the same category may have different sizes. Second, spectral values differ significantly between the seven categories. Generally, LS shantytowns are represented as gray pixels, while CM and ID buildings often consist of colored pixels. Some categories of buildings (e.g. AD buildings)