ISPRS Journal of Photogrammetry and Remote Sensing 129 (2017) 86–102

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Robust point cloud classification based on multi-level semantic relationships for urban scenes

Qing Zhu [a,b], Yuan Li [c], Han Hu [a,c,*], Bo Wu [c]

[a] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Gaoxin West District, Chengdu, China
[b] Collaborative Innovation Center for Geospatial Technology, 129 Luoyu Road, Wuhan, China
[c] Department of Land Surveying and Geo-Informatics, The Polytechnic University of Hong Kong, Hum Hom, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 23 November 2016
Received in revised form 27 April 2017
Accepted 27 April 2017
Available online 4 May 2017

Keywords:
Point clouds
3D semantics
Classification
Markov random field

ABSTRACT

The semantic classification of point clouds is a fundamental part of three-dimensional urban reconstruction. For datasets with high spatial resolution but significantly more noises, a general trend is to exploit more contexture information to surmount the decrease of discrimination of features for classification. However, previous works on adoption of contexture information are either too restrictive or only in a small region and in this paper, we propose a point cloud classification method based on multi-level semantic relationships, including point–homogeneity, supervoxel–adjacency and class–knowledge constraints, which is more versatile and incrementally propagate the classification cues from individual points to the object level and formulate them as a graphical model. The point–homogeneity constraint clusters points with similar geometric and radiometric properties into regular-shaped supervoxels that correspond to the vertices in the graphical model. The supervoxel–adjacency constraint contributes to the pairwise interactions by providing explicit adjacent relationships between supervoxels. The class–knowledge constraint operates at the object level based on semantic rules, guaranteeing the classification correctness of supervoxel clusters at that level. International Society of Photogrammetry and Remote Sensing (ISPRS) benchmark tests have shown that the proposed method achieves state-of-the-art performance with an average per-area completeness and correctness of 93.88% and 95.78%, respectively. The evaluation of classification of photogrammetric point clouds and DSM generated from aerial imagery confirms the method's reliability in several challenging urban scenes.

1. Introduction

Automatic three-dimensional (3D) city modeling has generated significant attention in the urban planning, analysis and design community. Despite the procedural approach (Dang et al., 2015; Esri, 2016; Vanegas et al., 2010), which uses predefined rules/-grammar and two-dimensional (2D) footprints to generate detailed 3D models, considerable efforts have been devoted to automatic reconstruction from point clouds (Lafarge and Mallet, 2012; Poullis, 2013; Xiong et al., 2015; Zhou and Neumann, 2010). Airborne laser scanning (ALS) is an important source of massive point clouds. Another important source is dense image matching (DIM), especially DIM using oblique images through multi-view stereo (MVS) pipelines (Furukawa and Ponce, 2010; McClune et al., 2016; Vu et al., 2012), which is quite popular in the field of photogrammetry currently. However, except for the generation of textured triangulated meshes, the automatic generation of 3D polygonal models remains an open problem that is being actively researched (Musialski et al., 2013). Recent advances in automatic urban reconstruction have revealed that enriching the raw point clouds or meshes with semantic segments and then reconstructing each segment, is an effective and scalable paradigm for large-scale reconstruction (Lafarge and Mallet, 2012; Poullis, 2013; Verdie et al., 2015).

However, the semantic segmentation or classification of point clouds, the focus of this paper, is considered non-trivial work in complex urban scenes (Bláha et al., 2016). The two cornerstones of classification are discriminative features and proper classifiers, both of which are generally obtained locally (e.g., a point and its neighborhood) (Chehata et al., 2009; Hackel et al., 2016), and only pairwise interactions (Niemeyer et al., 2014; Weinmann et al., 2015) are considered. However, due to the obvious defects of point

* Corresponding author at: Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Gaoxin West District, Chengdu, China.
E-mail address: huhan8807@gmail.com (H. Hu).

http://dx.doi.org/10.1016/j.isprsjprs.2017.04.022
0924-2716/

cloud data (e.g., noise, loss of sharp features and outliers), such methods are not resilient and thus require extensive drudgery in the form of manual quality control, especially for photogrammetric point clouds (Hu et al., 2016; Nex and Gerke, 2014). Compared to Light Detection and Ranging (LiDAR) point clouds, the photogrammetric are generally more noise-laden, which will dramatically decrease the discriminations of features derived from small local regions and consequently lead to the failure of classification. To achieve robust classification of point clouds, larger context range must be incorporated into the workflow to surmount the noise. In fact, the exploitation of larger context is a trend of scene classification. For example, by plane segmentation (Xu et al., 2014; Zhang and Lin, 2012) or second or higher order Markov random field (MRF) or conditional random fields (CRF) (Niemeyer et al., 2014, 2016; Lafarge and Mallet, 2012; Sengupta and Sturgess, 2015).

However, the adoption of contexture information is either too restrictive that requires perfect segmentation of planes (Zhang and Lin, 2012) or only involves a small local region through point-level interactions (Niemeyer et al., 2014; Lafarge and Mallet, 2012). Therefore, we propose a point cloud classification method that propagates the classification cues from a single point to the object level using flexible multi-level semantic relationships based on an intermediate representation of point clouds — the "supervoxel". The "supervoxel" in this paper is an extension of "superpixel" (Ren and Malik, 2003; Achanta et al., 2012) from 2D to 3D, but unlike the in the enumerative space of a 2D image, the "supervoxel" refers to a fixed-size cluster of unorganized points generated through space partitioning, and the points in each cluster maintain the original geometries individually but together constitute a regular shape. The proposed method involves three constraints derived from different entity levels. (1) the point–homogeneity constraint represents the semantic relationships between points and clusters homogenous points into over-segmented supervoxels designed to not cross object boundaries. (2) the supervoxel–adjacency constraint encodes pairwise interactions between supervoxels. (3) the class–knowledge constraint represents the global relationships of supervoxels at the object scale. These three relationships can be modeled using the unary, pairwise and high-order cliques (Li, 2009) in a MRF, and a two-step inference strategy is adopted to solve the labels of each supervoxel.

The remainder of this paper is organized as follows. Section 2 provides a brief literature review of the existing point classification methods. In Section 3, the classification method using multi-level semantic relationships is demonstrated in detail. The performance of the proposed methods is then evaluated and analyzed in Section 4, using both the ISPRS benchmark dataset (Rottensteiner et al., 2012) and photogrammetric point clouds derived from a penta-view multiple camera system (Petrie, 2009). The concluding remarks and future works are presented in Section 5.

2. Related works

According to the type of entity used for classification, existing methods can be categorized as point- or object/segment-based (Gerke and Xiao, 2014; Zhou et al., 2012). Below, we briefly review previous methods and demonstrate the rationale for the proposed method.

2.1. Point-based methods

Point-based methods generally extract point-wise features locally from the neighborhood defined by a sphere or cylinder, and then supervised or unsupervised classifiers are used. Therefore, such methods usually focus on the selection of discriminative features and effective classifiers. For instance, (Lodha et al., 2007) merged airborne LiDAR with images to extract more discriminative features, including geometric features from LiDAR and radiometric features from images. Then, based on these features, the points were divided into four classes with AdaBoost (Chehata et al., 2009). In addition to the geometric features from points and radiometric features from images, more sophisticated features are also used. For example, the full waveform LiDAR provides useful information for feature extractions (Mallet et al., 2008), spectral information within the feature selection framework shows promising results (Guo et al., 2011; Mallet et al., 2011) and hierarchical features exhibit superior performance in large-scale urban environments (Hackel et al., 2016). With regard to classifiers, despite the boosting method mentioned above, other popular methods such as Random Forests (RFs) (Breiman, 2001; Gislason et al., 2006) and support vector machines (SVMs) (Mountrakis et al., 2011) are also used for point cloud classification.

In the abovementioned methods, the points are labeled individually in the feature space without considering relationships, which often leads to discontinuities in the classification results. To avoid this, other point-based methods take advantage of contextual information. This type of semantic relationship at the point level leads to the use of graphical models, such as MRF or CRF (Kumar and Hebert, 2006). For instance, (Lafarge and Mallet, 2012) proposed an unsupervised method with an MRF framework, where the Potts model (Li, 2009) is introduced to define the pairwise interactions between neighboring points, with discriminative features from each point used to compute a potential classification result. Then, a graph cut-based algorithm (Boykov et al., 2001) was used to quickly reach an approximate solution close to the global optimum of energy. (Niemeyer et al., 2014) integrated an RF classifier into a CRF framework, where the unary and pairwise potentials of CRF were based on probabilities computed by RF.

Although these methods benefit from using contextual information, their effects have been very limited because they only consider the coherences between points within a small neighborhood. This limitation renders point-based methods less resilient to issues of data quality—such as noises and density anisotropy—such that these methods can generally only be applied to accurate point clouds (e.g., LiDAR), and require other ancillary datasets to extract discriminative features. As such, a great deal of manual parameter tuning and interactive post-processing work is required to refine the results. Furthermore, in cases of large-scale urban scenes, a typical site contains millions or more points, such that even a state-of-the-art inference method is challenged by graphical models with only pairwise interactions.

2.2. Object-based methods

Object-based methods choose a point cluster in which points share homogeneous properties as the entity to be classified. The features are generally extracted from the points first, and then split into different clusters, from which more discriminative features are extracted. The clusters are then classified based on these object-based features using a proper classifier, such as a probability distribution function (PDF) (Poullis, 2013), SVM (Zhang and Lin, 2012), RM or MRF (Gerke and Xiao, 2014). When compared with the point-based method, the major difference lies in the step for generating the clusters.

One strategy for generating clusters is to make each cluster contain as many homogeneous points (points that have similar colors, normals, curvatures, etc.) as possible through a segmentation process, so that each segment corresponds to a certain component of the objects, such as a façade or a roof. Based on geometric features, (Xu et al., 2014) segmented the point cloud into planar and irregular segments using surface growing (an instantiation of region