ISPRS Journal of Photogrammetry and Remote Sensing 87 (2014) 152–165

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Contextual classification of lidar data and building object detection in urban areas

Joachim Niemeyer [a,*], Franz Rottensteiner [a], Uwe Soergel [b]

[a] Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Nienburger Str. 1, D-30167 Hannover, Germany
[b] Institute of Geodesy, Remote Sensing and Image Analysis, TU Darmstadt, Franziska-Braun-Str. 7, D-64287 Darmstadt, Germany

ARTICLE INFO

Article history:
Received 8 July 2013
Received in revised form 4 November 2013
Accepted 4 November 2013
Available online 7 December 2013

Keywords:
LIDAR
Point cloud
Classification
Urban
Contextual
Building
Detection

ABSTRACT

In this work we address the task of the contextual classification of an airborne LiDAR point cloud. For that purpose, we integrate a Random Forest classifier into a Conditional Random Field (CRF) framework. It is a flexible approach for obtaining a reliable classification result even in complex urban scenes. In this way, we benefit from the consideration of context on the one hand and from the opportunity to use a large amount of features on the other hand. Considering the interactions in our experiments increases the overall accuracy by 2%, though a larger improvement becomes apparent in the completeness and correctness of some of the seven classes discerned in our experiments. We compare the Random Forest approach to linear models for the computation of unary and pairwise potentials of the CRF, and investigate the relevance of different features for the LiDAR points as well as for the interaction of neighbouring points. In a second step, building objects are detected based on the classified point cloud. For that purpose, the CRF probabilities for the classes are plugged into a Markov Random Field as unary potentials, in which the pairwise potentials are based on a Potts model. The 2D binary building object masks are extracted and evaluated by the benchmark ISPRS Test Project on Urban Classification and 3D Building Reconstruction. The evaluation shows that the main buildings (larger than 50 m^2) can be detected very reliably with a correctness larger than 96% and a completeness of 100%.

1. Introduction

Automated urban object extraction from remotely sensed data is a very challenging task due to the complexity of urban scenes. There are different types of objects such as buildings, low vegetation, trees, fences, and cars, that can be found in a small local neighbourhood, which makes it difficult to extract them reliably. In order to handle this problem, research often focuses on the extraction of a single object type, i.e. buildings, roads, and trees; for overviews, cf. Mayer (2008) and Rottensteiner et al. (2012).

Airborne LiDAR (Light Detection And Ranging) is a particularly useful technology for the acquisition of elevation data, with applications such as the generation of digital terrain models (DTM) (Kraus and Pfeifer, 1998), data acquisition for forestry (Reitberger et al., 2009), or power line monitoring (McLaughlin, 2006). LiDAR data are also well-suited for automated object detection for the generation of 3D city models. Building extraction is a prominent application in this context; two recent examples are Huang et al. (2013) and Liu et al. (2013).

For many applications a basic step in LiDAR processing is a classification of the point cloud. Each 3D point in the irregularly distributed point cloud is assigned to a semantic object class. Due to the complexity of urban scenes this task is also difficult. It is the goal of this paper to present an approach for the classification of a LiDAR point cloud in urban areas without the use of image data providing spectral information. The only radiometric signal feature we have access to is the so-called intensity, which is a function of the amount of photons collected by the scanning device. After the classification, 2D building outlines are delivered from the labelled point cloud.

1.1. Related work

In recent years research mainly focused on the use of supervised statistical methods for classification in remote sensing because they are more flexible to handle variations in appearance of the objects to be extracted compared to model-based approaches. Besides generative classifiers modelling the joint distribution of the data and labels (Bishop, 2006), modern discriminative methods such as AdaBoost (Chan and Paelinckx, 2008), Support Vector Machines (SVM) (Mountrakis et al., 2011),

* Corresponding author. Tel.: +49 511 762 19387; fax: +49 511 762 2483.
E-mail addresses: niemeyer@ipi.uni-hannover.de (J. Niemeyer), rottensteiner@ipi.uni-hannover.de (F. Rottensteiner), soergel@geod.tu-darmstadt.de (U. Soergel).

0924-2716/$ - see front matter © 2013 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.isprsjprs.2013.11.001

and Random Forests (RF) (Breiman, 2001; Gislason et al., 2006) are used. They usually lead to simpler models and need fewer training data in relation to generative models. These classifiers are also applied to LiDAR processing tasks. For instance, Mallet (2010) used a point-based multi-class SVM for the classification of full-waveform (FW) LiDAR data, whereas Chehata et al. (2009) applied RF for that purpose. However, both approaches classify each point independently without considering the labels of its neighbourhood. This is a drawback leading to inhomogeneous results in complex scenes such as urban areas, as demonstrated for example in Niemeyer et al. (2011). The reason is the diversity of objects' appearances even within a single scene. Especially in urban areas roofs of different shapes and other challenging objects with many details occur, leading to overlapping distributions of features within each class. Shadows caused by other objects, missing data due to the objects' properties, and random errors in the sensor data aggravate this effect. As a consequence, purely local decisions become uncertain.

An improvement can be achieved by incorporating contextual information, which is an important cue for the classification of objects in complex scenes. Spatial dependencies between the object classes can be trained to improve the results, because some object classes are more likely to occur next to each other than others; for instance, it is more probable that cars are situated on a street than on grassland. A sound statistical model of context leads to undirected graphical models (Bishop, 2006) such as Markov Random Fields (MRF) (Geman and Geman, 1984). In an MRF, the class label of an object is statistically dependent on its neighbours, whereas the data of different objects are assumed to be conditionally independent (Li, 2009). Conditional Random Fields (CRF) (Kumar and Hebert, 2006) offer a more general model. They drop the assumption of conditional independence of the data of different objects, expressed in the model of the unary potentials linking the class labels to the observations, and the interaction between neighbouring objects is modelled to depend on both the labels and the data in the pairwise potentials. CRFs have become a standard technique for considering context in classification processes, in particular for image classification (Kumar and Hebert, 2006; Schindler, 2012). They are also becoming more and more popular in the fields of photogrammetry and remote sensing. Some exemplary applications are multi-temporal image classification (Hoberg et al., 2012), building detection in radar images (Wegner et al., 2011), and classification of façade images (Yang and Förstner, 2011).

Applications of the CRF framework differ in the way they model the potentials and in the definition of the graph structure. For the unary potentials, the probabilistic output of a discriminative classifier is frequently used. Examples include linear models (Kumar and Hebert, 2006) and RF (Schindler, 2012). For the pairwise potentials, most approaches use relatively simple models favouring identical labels at neighbouring sites by penalising label changes, such as the Potts model. The contrast-sensitive Potts model (Boykov and Jolly, 2001) has the same effect, but adapts the degree of penalisation related to the Euclidean distance of the feature vectors. Schindler (2012) carried out a comparison of these smoothing models applied to high resolution images. Although both methods perform rather well in the comparison, these simple models tend to over-smooth the results. Thus, a more complex model might improve the results at the cost of higher computational efforts in training and of having to provide fully labelled training images. In Niemeyer et al. (2011) this was shown for the classification of LiDAR data of urban areas. In this case, linear models were used for both the unary and the pairwise potentials. In the latter case they were based on a multi-class model for the joint probability of the class labels at neighbouring sites rather than on a binary model for the probability of the two labels being equal. Nowozin et al. (2011) use RF classifiers for both types of potentials, also using a multi-class model for the interactions. In their examples, the random field is constructed over a (radiometric or depth) image grid. The neighbourhood system on which the edges of the graphical model are defined may vary with the application, but the interactions are restricted to pairs of nodes. Lucchi et al. (2012) use a CRF based on structured SVM (SSVM), which includes an SVM model for the pairwise terms. In their case, the graphical model is built on segments (superpixels), which reduces the computational complexity compared to a pixel-based classification.

Lucchi et al. (2011) have doubted the contribution of CRF-like models for classification, showing that methods for classifying superpixels and applying global features can achieve a similar performance as CRF-based models in classification of standard data sets. Their discussion is limited to images and to CRF-based models involving neighbourhood terms that depend on the relative alignment of objects in an image. They also show the effects of global constraints based on the co-occurrence statistics of objects in a scene. We think that the type of geometrical pairwise model used in Lucchi et al. (2011) (*"sky should appear above grass"*) is not applicable to remote sensing images, because it requires the definition of an absolute reference direction (e.g. the vertical in images having a horizontal viewing direction). Of course, height differences are important features in the context of point cloud classification, but the relative alignment in planimetry follows a similar structure as in aerial images. The benefits of using global energy terms such as those based on co-occurrence statistics, also proposed in Ladický et al. (2013), would also seem to be doubtful for the classification of remotely sensed images. In the urban remote sensing case, we usually have a small set of objects which always occur in a scene together (e.g., roads, buildings, trees and cars), so that the global information about their co-occurrence would not seem to carry much discriminative power.

The first research on the context-based classification of *point cloud* labelling was carried out in the fields of robotics and mobile terrestrial laser scanning. Anguelov et al. (2005) proposed a classification of a terrestrial point cloud into four object classes with Associated Markov Networks (AMN), a subclass of MRF. Neighbouring points are assumed to belong to the same object class with high probability, which leads to an adaptive smoothing of the classification results. In order to reduce the number of graph nodes, ground points are eliminated based on thresholds before the actual classification. Munoz et al. (2008) also used point-based AMNs, but they extended the original isotropic model to an anisotropic one, in order to emphasise certain orientations of edges. This directional information enables a more accurate classification of objects like power lines. Rusu et al. (2009) were interested in labelling an indoor robot environment described by point clouds. For object detection points are classified using CRFs according to the geometric surface they belong to, such as cylinders or planes. They applied a point-wise classification method, representing every point as a node of the graphical model. Compared to our application they deal with few points (≈80,000), and they even reduce this data set by about 70% before the classification based on some restrictions concerning the objects' positions. Shapovalov et al. (2013) also classified point clouds in indoor scenes, building a graphical model on point cloud segments. They consider long-range dependencies by so-called structural links, also based on special directions such as the vertical, the direction to the sensor or the direction to the nearest wall. In an indoor scenario, walls can be detected using heuristics (Shapovalov et al., 2013). However, in the airborne case, the number of points on walls is usually relatively low. The classifications of points on walls might be one of the problems one would like to solve by a CRF-based model, and the vertical and the direction to the sensor are nearly coincident. CRF were also used by Lim and Suter (2007) for the point-wise classification of terrestrial LiDAR data. They coped with the computational