ISPRS Journal of Photogrammetry and Remote Sensing 126 (2017) 129–145

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Semantic segmentation of forest stands of pure species combining airborne lidar data and very high resolution multispectral imagery

Clément Dechesne *, Clément Mallet, Arnaud Le Bris, Valérie Gouet-Brunet

Univ. Paris-Est, LASTIG MATIS, IGN, ENSG, F-94160 Saint-Mande, France

ARTICLE INFO

Article history:
Received 24 August 2016
Received in revised form 14 February 2017
Accepted 17 February 2017

Keywords:
Lidar
Multispectral imagery
Fusion
Feature selection
Supervised classification
Energy minimization
Regularisation
Forest stand delineation
Tree species

ABSTRACT

Forest stands are the basic units for forest inventory and mapping. Stands are defined as large forested areas (e.g., ⩾2 ha) of homogeneous tree species composition and age. Their accurate delineation is usually performed by human operators through visual analysis of very high resolution (VHR) infra-red images. This task is tedious, highly time consuming, and should be automated for scalability and efficient updating purposes. In this paper, a method based on the fusion of airborne lidar data and VHR multispectral images is proposed for the automatic delineation of forest stands containing one dominant species (purity superior to 75%). This is the key preliminary task for forest land-cover database update. The multispectral images give information about the tree species whereas 3D lidar point clouds provide geometric information on the trees and allow their individual extraction. Multi-modal features are computed, both at pixel and object levels: the objects are individual trees extracted from lidar data. A supervised classification is then performed at the object level in order to coarsely discriminate the existing tree species in each area of interest. The classification results are further processed to obtain homogeneous areas with smooth borders by employing an energy minimum framework, where additional constraints are joined to form the energy function. The experimental results show that the proposed method provides very satisfactory results both in terms of stand labeling and delineation (overall accuracy ranges between 84% and 99%).

1. Introduction

1.1. Motivation

Fostering information extraction in forested areas from remote sensing data, in particular at the stand level, is driven by two main goals: statistical inventory and mapping. Forest stands are the basic units for subsequent analysis and can be defined in terms of tree species or tree maturity. From a remote sensing point of view, the delineation of the stands is a segmentation problem. In statistical national forest inventory (NFI), an automated and accurate tree segmentation is needed in order to extract tree level features (basal area, dominant tree height, etc., (Means et al., 2000; Kangas and Maltamo, 2006)). However, the tree level is not the only reliable level of analysis for forest studies, a larger scale (e.g., forest stands) is interesting in order to extract reliable and statistically meaningful features and to provide an input for multi-source statistical inventory. For land-cover mapping, this is highly helpful for forest database updating (Kim et al., 2009), whether the labels of interest are *vegetated areas* (e.g., *deciduous/evergreen/mixed/non-forested)*, or, more precisely, the tree species. Most of the time in national forestry inventory institutes, for reliability purposes, each area is manually interpreted by human operators with very high resolution (VHR) geospatial images focusing on the infra-red channel (Kangas and Maltamo, 2006). This work is extremely time consuming and subjective (Wulder et al., 2008). Furthermore, in many countries, the wide variety of tree species (e.g., >20) significantly complicates the problem. The design of an automatic procedure based on remote sensing data would fasten such process. Additionally, the standard manual delineation procedure only takes into account the species, and few characteristics (alternatively height, age, stem density or crown closure), while an automatic method could offer more flexibility and would allow to combine characteristics extracted from all complementary data sources.

* Corresponding author.
E-mail address: clement.dechesne@ign.fr (C. Dechesne).

http://dx.doi.org/10.1016/j.isprsjprs.2017.02.011
0924-2716/

1.2. State of the art

The use of remote sensing data for the automatic analysis of forests has been growing in the last 15 years, especially with the synergistic use of airborne laser scanning (ALS) and optical VHR imagery (multispectral imagery and hyperspectral imagery) (Torabzadeh et al., 2014; White et al., 2016). In the large amount of literature in the field, only few papers focus on the issue of stand segmentation or delineation. They can be categorized with regard to the type of data processed.

First, stand segmentation can be achieved with a single remote sensing source. A stand delineation technique using VHR airborne multispectral imagery is proposed in Leckie et al. (2003). The trees are extracted using a valley following approach and classified into 7 tree species (5 coniferous, 1 deciduous, and 1 non-specified) with a maximum likelihood classifier. A semi-automatic iterative clustering procedure is then introduced to generate the forest polygons.

A hierarchical and multi-scale approach for the identification of stands is adopted in Hernando et al. (2012). The data inputs were the 4 bands of an airborne 0.5 m orthoimage (Red, Green, Blue, and Near Infra-Red) allowing to derive the Normalised Difference Vegetation Index (NDVI). The stand mapping solution is based on the Object-Based Image Analysis concept. It is composed of two main phases in a cyclic process: first, segmentation, then classification. The first level consists in over-segmenting the area of interest and performing fine-grained land cover classification. The second level aims to transfer the vegetation type provided by a land cover geodatabase in the stand polygons, already retrieved from another segmentation procedure. The multi-scale analysis appears to have a significant benefit on the stand labeling but it is highly heuristic and requires a correct definition of the stand while we consider it is an interleaved problem.

A seminal stand mapping method using low density airborne lidar data is proposed in Koch et al. (2009). It is composed of several steps of feature extraction, creation and raster-based classification. Forest stands are created by grouping neighboring cells within each class. Then, only the stands with a pre-defined minimum size are accepted. Neighboring small areas of different forest types that do not reach the minimum size are merged together to an existing forest stand. The approach offers the advantage of detecting 15 forest types that match very well with the ground truth but to the detriment of simplicity: the flowchart has to be highly reconsidered to fit to other stand specifications. Additionally, the tree species discrimination is not addressed.

The forest stand delineation proposed in Sullivan et al. (2009) also uses low density airborne lidar still coupling an object-oriented image segmentation and a supervised classification procedure. Three features are computed and rasterized. The segmentation is performed using a region growing approach. Spatially adjacent pixels are grouped into homogeneous discrete image objects or regions. Then, a supervised discrimination of the segmented image is performed using a Battacharya classifier, in order to determine the maturity of the stands. The tree species are ignored and the procedure requires a careful inspection of the raw data both for feature generation and model training.

Following the work of Wulder et al. (2008) with IKONOS images, Quickbird-2 panchromatic images are used in Mora et al. (2010) to automatically delineate forest stands. A standard image segmentation technique is used and the novelty mainly lies on the fact that its initial parameters are optimized with respect to NFI protocols. They show that meaningful stand heights can be derived, which are a critical input for various modeled inventory attributes.

The method proposed in Eysn et al. (2012) aims to generate a forest mask (*forested area* label only) using low density airborne lidar. A Canopy Height Model (CHM) with a spatial resolution of 1 m is derived. The positions and heights of single trees are determined from the CHM using a local maximum filter, based on a moving window approach. Only detected positions with a CHM height superior to 3 m are considered. The crown radii are estimated using an empirical function. The three neighboring trees are connected using a Delaunay triangulation applied to the previously-detected tree position. The crown cover is then calculated using the crown areas of three neighboring trees and the area of their convex hull for each tree triple. The forest mask is derived from the canopy cover values. While this is not a genuine stand delineation method, this approach could be easily extended to a multi-class problem and enlightens the necessity of individual tree extraction even with limited point densities as a basis for the stand-level analysis.

A forest stand delineation also based on airborne lidar data is proposed in Wu et al. (2014). Three features are first directly extracted from the point cloud. A coarse forest stand delineation is then performed on the feature image using the unsupervised Mean-Shift algorithm, in order to obtain under-segmented raw forest stands. A forest mask is then applied to the segmented image in order to retrieve forest and non-forest raw stands. It may create some small isolated areas, iteratively merged to their most similar neighbor until their size is larger than a user-defined threshold in order to product big raw forest stands. They are then refined into finer level using a seeded region growing based on superpixels. The idea is to select several different superpixels in a raw forest stand and merge them. This method provides a coarse-to-fine segmentation with relatively large stands. The process was only applied on a small area of a forest in Finland, thus, general conclusions can not be drawn.

Secondly, several methods fusing various types of remote sensing data have also been developed. The analysis of the lidar and multispectral data is performed at three levels in Tiede et al. (2004), following a given hierarchical nomenclature of classes in forested environments. The first level represents small objects (single tree scale, individual trees or small groups of trees) that can be differentiated by spectral and structural characteristics using a rule-based classification. The second level corresponds to the stand level. It is built using the same classification process which summarizes forest development phases by referencing to small scale sub-objects at level 1. The third level is generated by merging objects of the same classified forest-development into larger spatial units. The multi-scale analysis offers the advantage of alleviating the standard issue of individual tree crown detection and proposing development stage labels. Nevertheless, the pipeline is highly heuristic, under-exploits lidar data and significant confusion between classes are reported.

The automatic segmentation process of forests in Diedershagen et al. (2004) is also supplied with lidar and VHR multispectral images. The idea is to divide the forests into higher and lower sections with lidar. An unsupervised classification process is applied to the two new images. The final stand delineation is achieved by segmenting the classification results with pre-defined thresholds. The segmentation results are improved using morphological operators such as opening and closing, which fill the gaps and holes at a specified extent. This method is efficient if the canopy structure is homogeneous and requires a strong knowledge on the area of interest. Since it is based on height information only, it cannot differentiate two stands of similar height but different species.

In Leppänen et al. (2008) a stand segmentation technique for a forest composed of *Scots Pine*, *Norway Spruce* and *Hardwood* is defined. A hierarchical segmentation on the Crown Height Model followed by a restricted iterative region growing approach is performed on images composed of rasterized lidar data and Colored