Remote Sensing of Environment xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Identifying the genus or species of individual trees using a three-wavelength airborne lidar system

Brindusa Cristina Budei[a], Benoît St-Onge[b,*], Chris Hopkinson[c], Félix-Antoine Audet[b]

[a] Environmental Sciences Institute, University of Quebec at Montreal, Canada
[b] Department of Geography, University of Quebec at Montreal, Canada
[c] Department of Geography, University of Lethbridge, Canada

ARTICLE INFO

Keywords:
Tree
Species
Genus
Multispectral
Lidar
Intensity
Identification
Random forest
Titan

ABSTRACT

The identification of individual tree species is crucial for forest inventory, carbon stocks assessment as well as for habitat and ecosystem change studies. Previous studies have shown the potential of using both the geometrical information (proportions, shape of the crown profile, etc.) and return intensities of airborne laser scanning (ALS) point clouds to identify trees. So far, single wavelength (1064 nm or 1550 nm) ALS systems have been the most common. Teledyne Optech Inc. (Vaughan, Canada) introduced the Titan multispectral airborne lidar, the first three-wavelength system, equipped with 1550 nm, 1064 nm and 532 nm lasers. The objective of this study was to assess the accuracy of this discrete return multispectral lidar (MSL) for identifying the species of single trees, compared to standard discrete single wavelength lidar. Two distinct datasets were acquired in the Toronto region, Canada. The point clouds of single trees were extracted and 3D and intensity classification features were computed. New intensity features were developed for this purpose, such as lidar NDVIs (normalized difference vegetation index). The species of each tree were identified with a random forest classifier using features calculated from 1) each channel separately and 2) all channels. A stringent selection strategy was employed to reduce the number of features down to 5–9, from 99–142. Trees were classified either as broadleaved (BL) or needleleaf (NL), by genus (4–7 classes), or by species (10 classes). When using all channels, the classification accuracy surpassed what was achieved with single channels, but this advantage was only significant when the number of classes was high, i.e. in the case of seven genera, or ten species. Using MSL data, the out-of-bag error was 3–5% for the BL-NL classification, 13% and 20% for respectively four and seven genera, and 24% for ten species. In the latter case, the best single-channel classification (based on 1550 nm data) resulted in an error of 35%. In the MSL classification, the most useful features were the NDVIs (Normalized Difference Vegetation Index, based on the intensity of infrared and green channels) and the 1550 nm intensities. We therefore conclude that species identification accuracy can be improved by using a three-wavelength ALS such as the Titan system compared to single channel ALS systems, especially when tree species diversity is fairly large (seven classes or more).

1. Introduction

The identification of tree species is critical for forest inventory (Felbermeier et al., 2010), carbon stocks assessment (Jenkins et al., 2003) as well as for habitat and ecosystem change studies (Bradbury et al., 2005; Vastaranta et al., 2014). While the advantages of precise species data for the latter studies are self-evident, it should be underlined that species-specific allometric models for predicting timber volume or above-ground biomass (e.g. as a function of diameter at breast height and height) are more accurate than more general ones (Lambert et al., 2005; Tompalski et al., 2014). For decades, researchers have sought automated remote sensing methods for identifying tree species to replace photo-interpretation, a traditional approach (e.g., in North America and Europe) considered as time-consuming and subjective. Interpreters use three-dimensional appearance (3D), brightness and color cues to recognize species on aerial photography (Leckie et al., 1998; Eid et al., 2004), thus exploiting the architectural and reflectance differences between species. In boreal and temperate forests, important variations in structural characteristics indeed exist between species. The overall crown shape of most spruces (*Picea* sp.) and firs (*Abies* sp.), for example, is pointy and vertically elongated, compared to the roundish crowns of most broadleaved (BL) trees (such as *Acer*, *Betula*, *Populus*). Among needleleaf (NL) trees, branching patterns may differ. For example, white pines (*Pinus strobus*) and Norway spruces (*Picea abies*) seen from above exhibit a star-shaped outline, while white spruces (*Picea glauca*) and American larch (*Larix laricina*) have a smoother and more compact architecture. At least some of these 3D characteristics can be measured using

---

* Corresponding author.
E-mail address: st-onge.benoit@uqam.ca (B. St-Onge).

http://dx.doi.org/10.1016/j.rse.2017.09.037
Received 4 March 2016; Received in revised form 20 September 2017; Accepted 28 September 2017

Please cite this article as: Budei, B.C., Remote Sensing of Environment (2017), http://dx.doi.org/10.1016/j.rse.2017.09.037

airborne laser scanning (ALS) and used to identify species, by computing features, such as the height distribution of laser returns, features extracted using alpha shapes, or 3D textural properties (Vauhkonen et al., 2009; Li et al., 2013). When using single channel ALS (usually 1064 nm or 1550 nm) for basic species identification tasks, such as distinguishing between NL and BL trees or between just two NL species, the achieved accuracies were as high as 95% (Holmgren and Persson, 2004). Separating BL species, or distinguishing between a much larger number of species is however more challenging (Brandtberg, 2007; Kim et al., 2011). Some studies have tried to improve the identification accuracies by using full-waveform ALS features, such as pulse echo width, waveform amplitude, waveform energy, number of echoes, etc. (review: Koenig and Höfle, 2016). Accuracy improvements in species classification by about 6% (Vaughn et al., 2012) or 11% (Yu et al., 2014) were achieved by including full-waveform features, compared to identification based on discrete-return only.

The reflective properties of different species vary according to biochemical attributes of the foliage such as pigment concentration and nitrogen content (Ustin et al., 2009), structural attributes (leaf and stem structures, e.g., LAI and LAD), water content (Asner, 1998) and stem reflectance (Asner et al., 2014). For example, compared to NL species, boreal or temperate BL deciduous trees have leaves with a high growth rate during the short warm season and, consequently, have a large photosynthetic capacity and high concentrations of chlorophyll and nitrogen (Wright et al., 2004). To some degree, these differences can be discriminated based on their spectral signature.

Some limited improvement in tree species identification had been achieved by adding intensity features from the monospectral ALS (Ørka et al., 2009; Korpela et al., 2010b). More recently, a larger number of feature types were derived in a systematic way from 3D and intensity information, with improved results for species classification (Lin and Hyyppä, 2016). In general, approaches for capturing spectral radiometric data in addition to 3D information are comprised of a) using image matching to extract color photogrammetric point clouds from multispectral multi-view airborne images, b) combining single channel ALS to multi- or hyperspectral imagery, or c) combining airborne scanning lasers having different wavelengths. The first approach is still quite recent, but was used for example to identify three different tree boreal species with an accuracy of 89% (St-Onge et al., 2015). The second approach, which proceeds by extracting spectral information at the pixel locations corresponding to ALS first returns, either from multispectral (3–4 spectral bands) imagery (Persson et al., 2004; Holmgren et al., 2008; Ørka et al., 2012), or hyperspectral imagery (Dalponte et al., 2012; review: Ghosh et al., 2014), has brought improvements in the accuracy of species identification. However, coloring ALS points with image intensities requires joining two data sets usually obtained through distinct aerial surveys. Moreover, it is hindered by the complex directional reflectance anisotropy related to the changing sun-object-sensor geometry that influences intensities within and among images (Heikkinen et al., 2011). Considering these limitations, an emerging third approach was developed by using lidar sensors that transmit and receive at several wavelengths (Wang et al., 2013; Hopkinson et al., 2016). In addition to the increased density of the 3D component, the intensity of returns is measured in two or more wavelengths, allowing the creation of spectral signatures potentially useful for species recognition.

ALS intensities depend on the power of the backscattered laser pulses measured by the sensors. Instantaneous received power, explained by the radar equation (Jelalian, 1992; Wagner et al., 2006; Roncat et al., 2014), is determined by acquisition parameters (beam width, aperture, range, etc.) and the effective backscatter cross-section (σ, in m^2) of the reflecting object:

$$\sigma = \frac{4\pi}{\Omega}\rho A \tag{1}$$

where $4\pi/\Omega$ is the scattering angle of the object relative to an isotropic scatterer, ρ its reflectance, and A the object's area within a footprint (silhouette area). A is logically related to leaf area index, and $4\pi/\Omega$ to leaf angle distribution (Asner, 1998), two characteristics that vary between species, but that are wavelength independent. ρ however varies with wavelength and can theoretically enhance species discrimination if multispectral intensity data are acquired. Intensity values are obtained through a process in which the received power at a given time is converted to digital values on an arbitrary scale. In the case of discrete return laser scanners, intensities should be approximately proportional to the power of the received energy at the instant a return is triggered. However, the methods by which intensity values are generated by a given sensor are sometimes proprietary (e.g., for Teledyne Optech sensors) and not disclosed by the ALS system vendors, making it difficult to know exactly how received power is translated to digital values in these cases. It is nevertheless expected that the intensities will be correlated to σ, i.e. at least partly influenced by the values of ρ, which is itself species dependent.

Using ALS multispectral intensities rather than the spectral signature of passive sensors has certain theoretical advantages. The measurements of ALS intensity have the advantage of being independent from external illumination conditions. Because they are all done using the same “hot spot” geometry, the intensity data is not influenced by variable shadowing, leading to lesser variations compared to the aerial imaging case (Woodhouse et al., 2011). In addition, tree and ground signals cannot be separated in passive sensors' measurements of reflectance, whereas the association of the 3D and intensity data in ALS allows this discrimination by imposing a height threshold. Moreover, normalization of the radiometric variations caused by range difference in ALS data can be corrected to some extent for power attenuation due to the travel distance of pulses and their reflection (Korpela et al., 2010a). Several strategies for more advanced radiometric normalization have also been proposed (Yan and Shaker, 2014; review: Kashani et al., 2015).

The first attempts at using multispectral ALS data for land use, or vegetation classification usually had to rely on multiple airborne surveys, each using a single wavelength system. Wang et al. (2013) acquired two wavelength full-waveform ALS data in two separate flights using Optech ALTM Pegasus HD400 (1064 nm) and Riegl LMS-Q680i (1550 nm), and tested their performance in land cover classification. They highlighted the possibility of soil-vegetation separation, and the importance of moisture and physiological information of vegetation for species retrieval. Recently, Hopkinson et al. (2016) compared variation in intensity between discrete ALS point clouds acquired at three different wavelength, each with a different sensor on independent flights (Teledyne Optech's Aquarius - 532 nm, Gemini - 1064 nm, and Orion - 1550 nm). They found that the intensity-based foliage characterization was different for each sensor and was associated with both the sensor's wavelength and the survey sampling characteristics (flight altitude, system settings, etc.). Variations in the latter factors however makes it difficult to combine the intensities of different monospectral ALS to classify land cover or identify tree species using wavelength ratios as NDVI.

The operational advantages of integrating lasers of different wavelengths into a single system led to the development of multispectral laser (MSL) simulations and system prototypes. Some studies were conducted to analyze the variations of NDVI and photochemical reflectance index (PRI) along canopy profiles using a four wavelength airborne MSL (531, 550, 670 or 690, and 780 nm) simulated on virtual forest stands (Morsdorf et al., 2009) or tested in laboratory over living trees (Woodhouse et al., 2011). They demonstrated the possibility of capturing leaf-level physiological variations along vertical profiles and spatially distinguishing the photosynthetic active elements from bark material. Furthermore, different prototypes using a supercontinuum laser source were tested in laboratory. Chen et al. (2010) tested such a MSL prototype for NDVI estimation and collected range and intensity data at 600 nm and 800 nm. In another experiment, a full waveform hyperspectral terrestrial lidar using a supercontinuum laser was developed by Hakala et al. (2012) for the measurement of vegetation features at eight different wavelengths (542, 606, 672, 707, 740, 775, 878, and 981 nm). Nevalainen et al. (2014) used the same hyperspectral terrestrial lidar to test 27 vegetation indices and their relation with the chlorophyll amount. Furthermore Vauhkonen et al. (2013) tested this device for separating pines from spruces using intensity features and NDVI. For leaf nitrogen estimation and for different material classification, Gong et al. (2015) and Wei et al. (2012) tested a MSL with four wavelengths in the visible and infrared spectra (556, 670, 700, and 780 nm) transmitted from four semiconductor laser diodes.