INDIVIDUAL TREE SPECIES CLASSIFICATION USING STRUCTURAL FEATURES FROM HIGH DENSITY AIRBORNE LIDAR DATA

Jili Li, Baoxin Hu, Gunho Sohn, and Linhai Jing

Department of Earth and Space Science and Engineering, York University,
4700 Keele St., Toronto, Ontario, Canada, M3J 1P3
Email: ljili@yorku.ca

1. INTRODUCTION

Forest species are critical information for forest inventory and management. Conventionally, automatic species classification in forest is mainly conducted based on the spectral information in remote sensing imagery such as multispectral satellite imagery (e.g. [1], [2]). However, due to atmospheric conditions, illumination geometry, and optical properties of trees [3], spectral information of forest canopies is insufficient for individual tree species classification. For brevity, hereafter, we refer to individual tree species classification as species classification. Since different species have different structural properties, the latter are expectedly able to facilitate improving species classification. Airborne LiDAR (Light Detection and Ranging) technology can help to derive individual tree structural features and utilize them in species classification.

Current studies on species classification have been successfully carried out in leaf-on forests [4]. Although conifers and deciduous trees can be successfully classified using airborne LiDAR data (e.g. [5]-[7]), species within conifers or deciduous trees were less explored. This is primarily due to the usage of low density LiDAR data (1-10 points/m^2) which cannot fully characterize individual tree structures. Second, existing species classification methods are mostly limited by the selection of structural feature variables derived from LiDAR data, which is typically realized by the Linear Discrimination Analysis (LDA) or linear regression models [8]. In the feature selection methods, some variables important for specific species may be ignored. Moreover, species classification accuracy is greatly dependant on the forest structures and species combination of test sites. The latest studies in North America reported species classification accuracies of 73.1% [9] and 68% [10], indicating that species classification accuracy together with robustness can be considerably improved.

In this study, high density airborne LiDAR data (23 points/m^2) are used to fully exploit tree structure features for species classification within conifers or deciduous trees. The methods for deriving structural variables from LiDAR data are improved for high density data, and several new structural variables are constructed. Moreover, the Principal Component Analysis (PCA) method [11] is applied to optimize variable selection, and consequently some feature information which may be ignored by LDA can be utilized. A decision tree classification method [11] is developed and tested for automatically classifying species from selected feature variables.

2. METHODOLOGY

2.1. Feature extraction

Different species tend to have different crown shape and foliage distribution. In order to characterize the latter, two feature groups are extracted from each individual tree point cloud segment as listed in Table 1. To characterize crown shape, two geometric models are utilized to derive feature variables representing sharpness, symmetry and volume of individual trees. To characterize foliage distribution, variables are extracted by examining vertical profiles of individual tree, and are represent as different percentiles at certain intervals. Methods can refer to our previous study [12]. Moreover, in this study, the techniques of deriving variables Hp, Pp, and Cp are further improved to finer vertical percentiles (i.e. 5% height percentiles as described in Table 1.) because of the increased spatial resolution of LiDAR data. In addition, several 3 dimensional texture variables are newly developed in this study. To do that, a voxel space is firstly defined for each individual tree segment and 3 dimensional volume data which represent point density in small voxels are created from the voxel space. The variables of 3 dimensional textures in Table.1 stand for the Haralick texture features [13] generated from the extended gray-level co-occurrence matrix using the volume data.

2.2. Feature selection and classification

Firstly, to reduce the dimensional of variable groups Hp, Pp, and Cp (Table. 1) without losing important feature information, PCA is applied to the three groups respectively to obtain the first and second principal components for each group. Secondly, multivariable correlation matrix of the derived principal components together with other variables in Table.1 is calculated, in order to uncover the highly correlated variables which will not be selected for species classification. Feature variables with small correlations are selected as the inputs of decision tree algorithm. Furthermore, the proposed decision tree classifier is mainly designed to classify the species within conifers and within deciduous trees after the separation of the two categories. Decision rules are automatically derived from the training data of individual trees. Finally, the classification results are validated by field data.

3. EXPERIMENT

The discrete airborne LiDAR data were generated from small footprint full-waveform data collected in August, 2009. The density of the discrete data is 23 points/m^2 in average. The test site consists of 6 forest research plots located in the east of Sault Ste. Marie, Ontario, Canada. The sites are boreal forest including both mature coniferous and deciduous stands. Five dominant species jack pine (Pinus banksiana), white pine (Pinus strobus), white birch (Betula papyrifera), aspen (Populus tremuloides), and sugar maple (Acer saccharum) were selected from the 6 forest research plots. Field data were collected at the same time with the LiDAR data acquisition. 276 individual tree segments were isolated from LiDAR point clouds and used in this study. 79 variables were extracted from each segment firstly. After the application of PCA method, the number of variables was reduced to 25 and then 15 highly correlated variables were further removed based on correlation matrix. Half of the 276 trees were used for training of the decision tree classifier and the others for validation. The overall classification accuracy of the five species is 81.3%.

Table 1. Feature variables derived from LiDAR data

Feature groups	Feature variables	Description
Crown Shape (F_{CS})	FSH	Sharpness of tree top
	FSY	Symmetry of tree top
	FNV	Normalized volume
	FHD	Height over maximum crown diameter
Foliage Distribution (F_{FD})	Hp1, Hp2, … , Hp20	Height percentiles at every 5% point intervals
	Pp1, Pp2, … , Pp20	Point density percentiles at 5% height intervals
	Cp1, Cp2, … , Cp20	Crown area percentiles at 5% height intervals
	PRF, PRM, PRL	Proportion of first, middle, and last returns
	MEANI, STDI	Mean intensity, stand derivation of intensity of all returns
	T1, T2, … , T10	3D texture variables

4. CONCLUSION

The study demonstrated the usefulness of high density airborne LiDAR data for species classification in boreal forest of North American. In this study, a new 3 dimensional texture feature is developed and verified to be useful to characterize foliage distribution of individual trees. PCA method and correlation matrix are well applied to optimize the variable selection. Moreover, experimental results illustrate that decision tree algorithm is efficient for species classification. Comparing with the classification accuracy in similar boreal forest areas of Ontario reported years ago [14], the proposed methods in this study obtained a higher overall accuracy above 80%. Within the five species, paper birch and aspen are relatively hard to differentiate due to the similar crown shape and foliage distribution. Future research will be conducted on the integration of airborne LiDAR data and high-resolution multispectral imagery.

5. REFERENCES

[1] A. Carleer and E. Wolff, "Exploitation of very high resolution satellite data for tree species identification", Photogrammetric Engineering & Remote Sensing, vol. 70, pp. 135-140, 2004.

[2] D.A. Roberts, S.L. Roberts, S. Ogunjemiyo, J. Greenberg, S.Z. Dobrowski, and J. Chen, "Spectral and structural measures of northwest forest vegetation at leaf to landscape scales", *Ecosystems*, vol. 7, pp. 545-562, 2004.

[3] G. Guyot, D. Guyon, and J. Riom, "Factors affecting the spectral response of forest Canopies: a review", *Geocarto International*, vol. 4, pp. 3-18, 1989.

[4] J. Hyyppä, H. Hyyppä, D. Leckie, F. Gougeon, X. Yu, and M. Maltamo, "Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests", *International Journal of Remote Sensing*, vol. 29, pp. 1339-1366, 2008.

[5] J. Reitberger, P. Krzystek, and U.Stilla, "Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees", *International Journal of Remote Sensing*, vol. 29, pp. 1407-1431, 2008.

[6] X. Liang, J. Hyyppä, and L. Matikainen, "Deciduous-Coniferous tree classification using difference between first and last pulse laser signatures", *IAPRS Volume XXXVI, Part 3 / W52*, pp. 253-257, 2007.

[7] T. Moffiet, K. Mengersen, C. Witte, R. King and R. Denham, "Airborne laser scanning: exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 59, pp. 289-309, 2005.

[8] H.O. Ørka, E. Næsset, and O.M. Bollandsås, "Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data", *Remote Sensing of Environment*, vol. 113, pp. 1163-1174, 2009.

[9] S. Kim, R.J. McGaughey, H. Andersen, and G. Schreuder, "Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data", *Remote Sensing of Environment*, vol. 113, pp. 1575-1586, 2009.

[10] A. Suratno, C. Seielstad, and L. Queen, "Tree species identification in mixed coniferous forest using airborne laser scanning", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, pp. 683-693, 2009.

[11] W. Sun, J. Chen, and J. Li, "Decision tree and PCA-based fault diagnosis of rotating machinery", *Mechanical Systems and Signal Processing*, vol. 21, pp. 1300-1317, 2007.

[12] J. Li, B. Hu. G. Sohn, "A structural analysis for individual tree classification using airborne LiDAR data", *Proceeding of Canadian Symposium and Remote Sensing*, Lethbridge, Canada, June 2009.

[13] R.M. Haralick, "Statistical and structural approaches to texture", *Proc. IEEE*, vol. 67, pp. 786-804, 1979.

[14] I.D. Thompson, S.C. Maher, D.P. Rouillard, J.M. Fryxell, and J.A. Baker, "Accuracy of forest inventory mapping: Some implications for boreal forest management", *Forest Ecology and Management*, vol. 252, pp. 208-221, 2007.