Tree genera classification using airborne LiDAR data by ensemble methods

Connie Ko[1], Tarmo K. Remmel[2], Gunho Sohn[3], John R. Miller[4]

1 Earth & Space Science & Engineering, York University, cko@yorku.ca
2 Geography, York University, remmelt@yorku.ca
3 Earth & Space Science & Engineering, York University, gsohn@yorku.ca
4 Earth & Space Science & Engineering, York University, jrmiller@yorku.ca

Abstract

We propose an ensemble classification method for classifying tree genus by using LiDAR (Light Detection and Ranging) data. We have developed a set of descriptors (features) related to the geometric information given by the point cloud. The second set of features is derived from a more conventional method and is related to the vertical point distribution of the point cloud. We built two classifiers separately (geometric classifier and vertical profile classifier) using the two sets of features and then combine the classifiers for improving overall classification accuracy. Our study area is located north of Thessalon, Ontario, Canada and we are classify trees into three genera (pine, poplar and maple) within our study sites. Result show that the average classification accuracy for the geometric classifier is 88.0% and 88.8% for vertical profile classifier. When the classifiers are combined, the overall accuracy improved to 91.2%.

Background and Relevance

The use of aerial LiDAR in forestry applications has become increasingly popular for its ability to acquire 3D information and has proven successful in tree species/genera classification (Holmgren and Persson 2004; Brandtberg 2007; Holmgren et al., 2008; Kato et al., 2009; Ørka et al., 2009; Vauhkonen et al., 2009; Korpela et al., 2010 and Kim et al., 2011). Our first set of features are derived from the geometry of the LiDAR point distribution, this approach can be found in Kato et al. (2009) where the authors fit curved surfaces to the individual LiDAR tree crown and Vauhkonen et al. (2009) compute alpha shapes of the LiDAR tree crowns. Both methods derive features related to the outer shape of the tree crown, we further develop features that relate to the outer as well as inner geometry of the tree (branching levels). The second set of features is calculated from a more convention approach, examples of such an approach include Holmgren and Persson (2004); Brandtberg (2007); Ørka et al. (2009); Korpela et al. (2010) and Kim et al. (2011). These authors derived features from the vertical point profile reflected from the tree (or tree crown) and calculate statistical metrics that summarizes the point distribution within specific height percentiles or the entire profile. The advantage of the geometric features is that they can be easily related to the physical and biological implication of tree form, however they are usually more computationally expensive. Conversely, vertical profile features are computationally efficient but are less intuitive. This research takes advantage of the both perspectives and combines both classifiers to yield a better result.

Methods and Data

The study area is located north of Thessalon, about 75 km east of Sault Ste. Marie, Ontario, Canada. We have selected eight field sites in the area and ground validated 186 trees. We have identified white birch (*Betula papyrifera* Marsh.), balsam fir (*Abies balsamea* (L.)), maple (*Acer saccharum* Marsh.), red oak (*Quercus rubra* L.), jack pine (*Pinus banksiana* Lamb.), poplar (*Populus* temuloides), white pine (*Pinus strobus* L.) and white spruce (*Picea glauca* (Moench Voss)) at the field sites. 160 of the trees belong to three broader genera, pine, poplar and maple and therefore will be used for classification. LiDAR data was collected by a Riegl LMS-Q560 scanner, the flight altitude varies between 122 m to 250 m above ground level with point density approximately 40 pulses / m^2 with up to five returns per pulse.

The methods for deriving the geometric features are detailed in Ko et al. (2013) and Table 1 summarizes the six selected features for this research; Table 2 summarizes the 26 selected vertical profile features for this research.

Table 1.

No.	Descriptions for selected geometric features
1	Mean line segment lengths derived from point cloud divided by height of the tree
2	Mean line segment lengths multiplied by tree crown height to tree height ratio
3	Convex hull volume calculated from tree crown divided by total number of points
4	Mean orthogonal distances from each LiDAR point to the closest facet of the convex hull
5	After buffer each LiDAR point by a radius of 2% of the tree height, sum the overlapped volume of the spheres and divided by the total number of points
6	Ratio between tree crown height and tree height

We use Random Forests (Brieman 2001; Liaw and Wiener 2002) implemented in R (R Development Core Team 2013) for classification. 25% of the dataset is partitioned for training the classifiers and 75% of the data is partitioned for validation. This optimal partition is an experimental result from Ko et al. (2013). The classification is performed separately with two classifiers and then combined by the following strategy. Using the geometric classifier as a base classifier, we automatically filter out trees that are potentially misclassified. Then these trees are classified by the vertical profile classifier and classification results are compared to the initial result classified by the geometric classifier. A final decision is made by the classifier that obtains a larger margin from the prediction provided by Random Forests.

Table 2. Descriptions for selected vertical profile features

	First returns only	Single returns only	Last returns only
% of canopy return		F1	F2
% of return count at 10th percentile	F3	F4	F5
% of return count at 90th percentile	F6	F7	F8
Mean height of canopy return	F9		F10
SD of height	F11	F12	
SD height for canopy return	F13		F14
CV height for canopy return	F15	F16	
Kurtosis of variation height for canopy return		F17	F18
Skewness of variation height for canopy return		F19	F20
Mean intensity at 10th percentile			F21
Mean intensity at 90th percentile	F22	F23	
SD of intensity	F24		
CV intensity of canopy return			F25
Skewness of variation intensity of canopy return		F26	

SD= standard deviation; CV = coefficient of variation

Results

Table 3a shows the confusion matrix for geometric classifier; Table 3b shows the confusion matrix for vertical profile classifier. Table 4 shows the confusion matrix that demonstrates the performance of our ensemble classification. The results shown are based on using 75% of the data set, repeated 20 times, using (25%) as training samples, resulting 2400 tree samples in total for each confusion matrix.

Table 3a shows that the average classification accuracy for geometric classifier is 88.0% and 88.8% for vertical profile classifier from Table 3b. Both classifiers have the highest classification accuracy in classifying maple trees. Thus, the accuracies for classifying pine and poplar are lower, this is because point density distribution for pine and poplar are similar. However, we can still observe some accuracy discrepancies between the two classifiers, where largest difference is observed in the producer's accuracy for classifying pine. This indicates that there is a potential benefit to ensemble classification. By comparing the use of a geometric classifier alone with results obtained from Table 4 (where the geometric classifier is selected as base classifier), accuracies for all genera improve.

Table 3a. Confusion matrix for geometric classifier (average accuracy = 88.0%)

		Actual class			User's Accuracy (%)
		Pine	Poplar	Maple	
Predicted class	Pine	856	115	19	86.5
	Poplar	123	771	2	86.0
	Maple	27	1	486	94.6
Producer's Accuracy (%)		85.1	86.9	95.9	

Table 3b. Confusion matrix for vertical profile classifier (average accuracy = 88.3%)

		Actual class			User's Accuracy (%)
		Pine	Poplar	Maple	
Predicted class	Pine	906	132	12	86.3
	Poplar	87	736	7	88.7
	Maple	13	19	488	93.8
Producer's Accuracy (%)		90.1	83.0	96.3	

Table 4. Confusion matrix for ensemble classification (average accuracy = 91.2%)

		Actual class			User's Accuracy (%)
		Pine	Poplar	Maple	
Predicted class	Pine	903	94	5	90.1
	Poplar	86	786	3	89.8
	Maple	17	7	499	95.4
Producer's Accuracy (%)		89.8	88.6	98.4	

Conclusions

In this research, we applied ensemble methods that combine features derived from the geometry of LiDAR points reflected from individual trees with features derived from the vertical point distribution. Although geometric features have advantages over vertical profile features in terms of tying the close relationship with tree form, the advantages of vertical profile features should not be overlooked. Thus we combine both methods for improving classification accuracy. Table 3a and Table 3b shows that individual classifiers make different decisions and the differences indicate there is a potential for improving accuracy after combining the classifiers. By combining the decisions made by the two classifiers, the classification accuracy improved to 91.2% (Table 4) if the geometric classifier is being used as the base classifier. Since the original accuracies (with a single classifier) are already very high, the marginal improvement that has been made represents an improvement that is difficult to attain by traditional methods.

References

Brandtberg, T. (2007). Classifying individual tree species under leaf-off and leaf-on conditions using airborne LIDAR. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(5), 325–340.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

Holmgren, J., and Persson, Å. (2004). Identifying species of individual trees using airborne laser scanning. *Remote Sensing of Environment*, 90(4), 415-423.

Holmgren, J., Persson, Å, and Söderman U. (2008). Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *International Journal of Remote Sensing*, 29(5), 1537-1552.

Kato, A., Moskal, L.M., Schiess, P., Swanson, M.E., Calhoun, D., and Stuetzle, W. (2009). Capturing tree crown formation through implicit surface reconstruction using airborne lidar data, Remote Sensing of Environment, Vol. 113, No. 6, pp. 1148-1162.

Kim, S., Mcgaughey, R.J., Andersen, H., and Schreuder, G. (2009). Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sensing of Environment*, 113(8), 1575-1586.

Ko, C., Sohn, G., and Remmel, T. K. (2013). Tree genera classification with geometric features from high-density airborne LiDAR. *Canadian Journal of Remote Sensing*, 39(S1), S1-S13.

Korpela, I., Ørka, H. O., Maltamo, M., and Tokola, T. (2010). Tree species classification using airborne LiDAR—Effects of stand and tree parameters, downsizing of training set, intensity normalization and sensor type. *Silva Fenn*ica, 44(2), 319–339.

Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18-22.

Ørka, H.O., Næsset, E., and Bollandsås, O.M. (2009). Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sensing of Environment*, 113(6), 1163-1174.

R Development Core Team. (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, 2013. URL http://www.R-project.org/.

Vauhkonen, J., Tokola, T., Packalén, P., and Maltamo, M. (2009). Identification of Scandinavian commercial species of individual trees from airborne laser scanning data using alpha shape metrics. *Forest Science*, 55(1), 37-47.