A DECIDIOUS-CONIFEROUS SINGLE TREE CLASSIFICATION AND INTERNAL STRUCTURE DERIVATION USING AIRBORNE LIDAR DATA

C. Ko [a,] *, G. Sohn [a], T. K. Remmel [b]

[a] Department of Earth and Space and Engineering, York University, Toronto, M3J 1P3, Canada - cko@yorku.ca
[a] Department of Earth and Space and Engineering, York University, Toronto, M3J 1P3, Canada - gsohn@yorku.ca
[b] Department of Geography, York University, Toronto, M3J 1P3, Canada - remmelt@yorku.ca

Commission III, WG 2

KEY WORDS: LiDAR, species classification, geometrical analysis, bole derivation, branch derivation

ABSTRACT:

This project has two main purposes; the first is to perform deciduous-coniferous classification for 65 trees by using the leaf-on single flight LiDAR data. It was done by looking at the geometrical properties of the crown shapes (spherical, conical or cylindrical), these shapes were developed by a rule-driven method Lindenmayer Systems (L systems). Two more parameters that are data driven (convex hull analysis and buffer analysis) were developed to further capture the geometrical differences between deciduous and coniferous trees. Proposed methods are scale independent and arithmetically simple, they were developed simply by looking at the geometrical differences between the two types of trees. The classification rate was cross-validated and trees are 85% - 88% correctly classified. The second part of the project is to derive the internal structures of the LiDAR tree according to the results obtained from the first part. Internal structures include bole and branches; the location and orientation of the bole was done by connecting the geographic centres of horizontal slices of the tree. The branches were derived by k-means clustering techniques, different types of trees will yield a different type of branching structures for better visualization.

1. INTRODUCTION

The taxonomic classification of vegetation, especially trees, has been a piece of useful information for many studies, but it is a challenging task because remotely sensed imagery that are typically obtained from a nadir (or off-nadir) viewing angle, provide little information about the internal structures of tree canopies. In many studies, tree classifications are done by human interpretation using aerial photos. The introduction of small footprint airborne LiDAR opened up many research possibilities for forest studies because of the capability of LiDAR to penetrate canopies vertically and revealed some of their internal structures, thus, providing geometrical information about tree crowns and boles. For that reason, it is logical to develop methodologies that include the internal structures of individual trees. We have selected five parameters that involve the consideration of the third dimension. The results shows that by incorporating three-dimensional geometric characterizations of tree crowns obtained from LiDAR point clouds, we are able classify trees into deciduous and coniferous categories.

A substantial quantity of previous work on studying forest parameters and the classification of tree species using small footprint LiDAR has been done by flying over the canopy during leaf-off conditions (Brandtberg et al., 2003; Liang et al., 2007). The advantages of using leaf off data are to get better internal branching structures or to get the difference between the first last pulse for deciduous-coniferous classification. However, leaf-on data provide accurate forest parameters such as height, crown size, crown shape, and tree-top shape (Holmgren and Persson, 2004; Holmgren et al., 2008 and Litkey et al., 2007). Also, Reitberger et al. (2008) show that it is better to use leaf on LiDAR data for deciduous crown representation. It would be ideal to have the two conditions for classification purposes but in many cases, data is available for either leaf-on or leaf-off condition. One of the goals of this paper is to classify trees as deciduous or coniferous with single leaf on flight data therefore benefit both conditions.

Some work that deals with tree crown shape or crown fitting involve fitting parabolic surfaces to the canopy height model (Persson et al., 2002; Holmgren and Persson, 2004), we believe that LiDAR point clouds not only provide us information about the surface shape, there are also useful information inside the crown that we should consider. Therefore, instead of using surface fitting, we try to develop models and rules that consist of internal point clouds (L system trees). Five parameters were used to develop a decision tree for classification purposes.

LiDAR applications in forestry have two main approaches, the first one is canopy height distribution approach and the second one is an individual tree-based approach (Hyyppä at el., 2008). The objective of this paper is to classify trees as either deciduous or coniferous by using three-dimensional tree crown geometry information obtained during a single leaf-on flight. The second part of the paper is to reconstruct the internal branching structures (bole and branches) by implementing k-means clustering techniques for improved visualization.

2. METHODS:

2.1 Classification of trees (deciduous and coniferous by geometry)

The purpose of this section is to use the geometrical properties of tree crowns for deciduous and coniferous classification. LiDAR data was collected on September 7, 2008 at State of

* Corresponding author.

Pennsylvania; United State of America, using a Riegl LMS-Q560 /LMS Q280i at an altitude of 121.9 m above ground level and speed of 23.15 m/s. The swath width is approximately 190 m and it covered about 233,000 m^2, the point density is twenty-one points per m^2. 65 individual tree crowns were separated manually and further investigated individually in this paper. The minimum crown height (difference between the highest point and the lowest point in each crown) is 3.5 m and the maximum crown height is 27.5 m, the mean height is 10.1 m. There are 27 trees classified as coniferous and 38 classified as deciduous.

2.1.1 Creation of L System tree – shape characteristics:

L systems was developed by Aristid Lindenmayer originally modelling cellular development, it was then applied to model plant growth (Prusinkiewicz and Lindenmayer, 1990). The purpose of creating L system trees is to study the relationship between crown shape and tree classification. L system is a language contains two elements, the axiom (one rule) is like a starting point of a tree and the production (a set of rules) is like the growing patterns of the tree. Tree-like structures were created based on the movement of user defined productions and each movement (each node and connectivity) were recorded. Three types of these tree-like structures are created with various rules. For this paper, they are generally described as 1. Spherical (Figure 1a), 2. Conical (Figure 1b), and 3. Cylindrical (Figure 1c) with each of their nodes recorded as point triplets (x, y, z). Each shape was produced in three point density levels (500, 2000, and 7000 points/crown respectively) to accommodate different sizes of the actual crown. Generated trees are volume scaled by the height of the crown and translated to the coordinate system of the real trees, each of the 65 trees were compared to the three generated crown shape and mean of the minimum Euclidean distance from each point on real tree to the L system tree is calculated and recorded. Figure 2a, 2b and 2c shows the fitting between one of the LiDAR trees compared to the three L system trees. These three shapes were chosen because deciduous trees were believed to be associated with spherical crowns and coniferous trees were believed to be associated with conical and cylindrical shaped crowns (Horn, 1971). A normalized model closeness between LiDAR points and generative models produced based on L systems for each tree species is measured as follows:

$$\text{Normalized spherical distance} = \left(\frac{D_s}{D_s + D_c + D_{cy}} \right) \tag{1}$$

$$\text{Normalized conical distance} = \left(\frac{D_c}{D_s + D_c + D_{cy}} \right) \tag{2}$$

$$\text{Normalized cylindrical distance} = \left(\frac{D_{cy}}{D_s + D_c + D_{cy}} \right) \tag{3}$$

where D_s = mean min. inter-point distance for spherical trees
D_c = mean min. inter-point distance for conical trees
D_{cy} = mean min. inter-point distance for cylindrical trees

and D_s, D_c and D_{cy} are calculated by:

$$\frac{\sum_{n=1}^{k} \min \left(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \right)}{k} \tag{4}$$

Where $P_i(x_i, y_i, z_i)$ be any point on the real tree; $P_j(x_j, y_j, z_j)$ be any point on the L system tree; k= number of points in real tree.

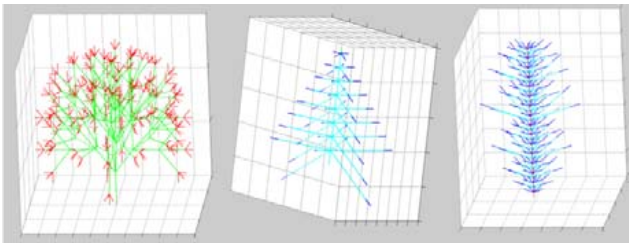


Figure1a Figure 1b Figure 1c

Figure 1a shows the structure created and described as spherical crown shape, Figure 1b shows the structure described as conical crown shape and Figure 1c shows the structure described as cylindrical crown shape

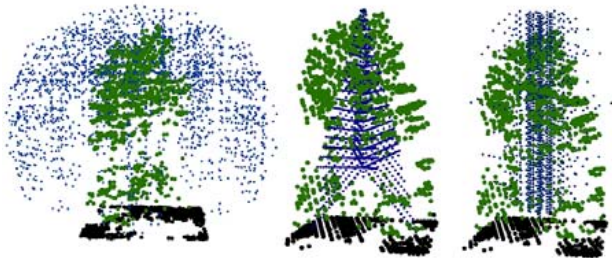


Figure 2a Figure 2b Figure 2c

Figure 2a shows the fitting between real tree LiDAR point data (green) and L system generated spherical crown (blue). Figure 2b shows the fitting between real LiDAR (green) and L system conical crown (blue). Figure 2c shows the fitting between real LiDAR (green) and L system cylindrical crown (blue)

2.1.2 Convex hull calculation – area to volume ratio:

For each tree crown, the convex hull was created, the volume of the convex hull was then calculated and the surface area of the convex hull is the summation of all triangular facets comprising the hull. This is done to study the spherical nature of the tree crown, the closer it is to 1.5, the more spherical the tree crown.

$$HullRatio = \frac{A_{hull}}{V_{hull}} \times \frac{h_{crown}}{2} \tag{5}$$

=3 for sphere

where V_{hull} = Volume of the convex hull of the tree crown
A_{hull} = Surface area of the convex hull
h_{crown} = height of the tree crown = maximum height – minimum height for the tree crown

The factor (height/2) was multiplied to correct the dimension between volume and area. Figure 3a and 3b shows visualization of the convex hull of one sample tree.

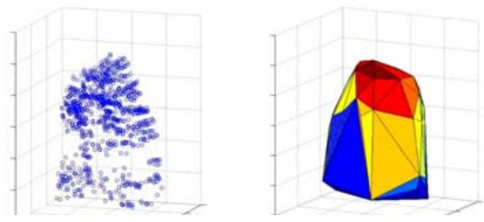


Figure 3a Figure 3b

Figure 3a shows LiDAR point data of a tree crown. Figure 3b: convex hull of the tree crown created

2.1.3 Buffer analysis – ratio between number of polygons and points: The buffer analysis was done to study the layering properties of coniferous trees, the individual branch layers are separated by a relatively large distance and also branches are flatly layered so that the trees can tolerate wind during winter (Smith and Brewer, 1994). Therefore, LiDAR points are usually clustered on the branching planes, such that when a tree is viewed from the side, layers of LiDAR points can usually be found in coniferous trees whereas for deciduous trees LiDAR points are more evenly distributed.

First, each tree is projected onto a vertical two dimensional plane, then, points are thinned out randomly so that each tree crown has about 200 points; this is to avoid over-aggregation of buffered polygons for high density trees later. Next, buffer polygons were created using buffer distance = 2% of the crown height, then, the ratios for all 65 trees were calculated as:

$$Ratio = \frac{number\ \ of\ \ polygon}{number\ \ of\ thinned\ \ out\ LiDAR\ \ po\ int\ s} \qquad (6)$$

Figure 4 shows the result from one of the sample trees. Lower ratios indicate the likelihood of being a conifer

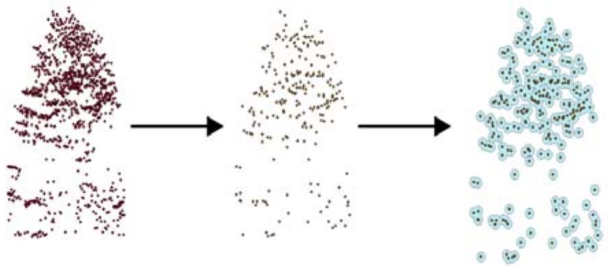


Figure 4. Result of a sample tree for calculating the ratio of buffered polygon to thinned out LiDAR points

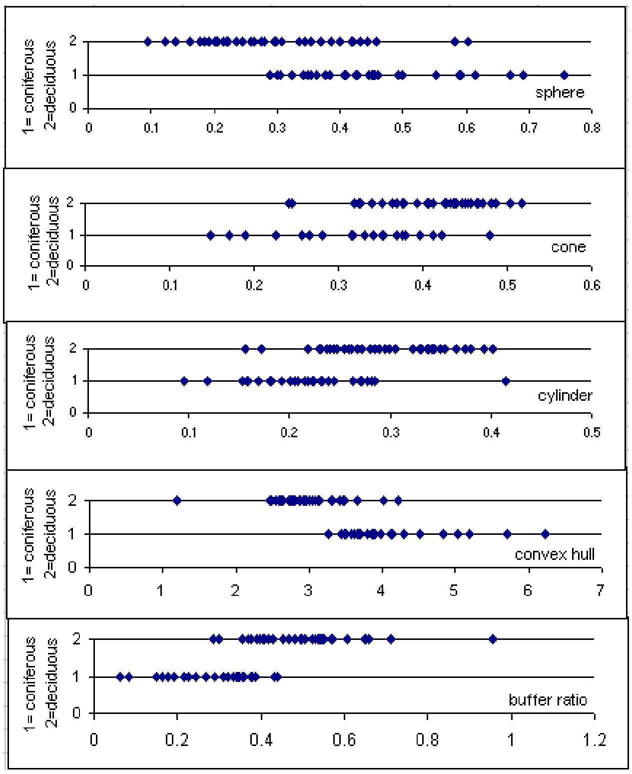


Figure 5. Scatter plots of types of trees versus normalized spherical distance; normalized conical distance; normalized cylindrical distance; convex hull volume area ratio and polygon to point ratio in buffer analysis

2.2 Decision tree classification

Figure 5 shows the results calculated from 2.1.1 to 2.1.3 in five different scatter plots, all the five variables shows there are some separations between coniferous and deciduous trees. However, all of the variables have some overlaps between coniferous and deciduous. Therefore, a tree classification was performed by using the above five parameters as decisions in an attempt to separate the 65 trees into two classes: deciduous and coniferous. The algorithm used (see Breiman et al. 1984) assumes that each split is binary. Then, by assigning random samples at each node, the deviances of each node can be calculated, the best nodes are chosen by maximizing the reduction of deviance in each split.

Let:

Decision 1 = Normalized spherical distance between L system and true trees
Decision 2 = Normalized conical distance between L system and true trees
Decision 3 = Normalized cylindrical distance between L system and true trees
Decision 4 = Corrected volume to surface area ratio of tree crown hull
Decision 5 = Ratio of buffered polygon to LiDAR points

The classification tree was pruned at the level with the minimum number of nodes (the most possible simple tree) when the cross validation (10-fold cross validation) error is within one standard error of the minimum error. Figure 6 shows the error rates for the two classification trees with increasing number of nodes for each tree. There were six nodes and four nodes for un-pruned classification tree using decisions 1-3 and 1-5 respectively. In the second classification tree (involve decisions 1-5), the first three decisions were automatically removed because they were relatively weak classifiers. From the results obtained in figure 6, decision tree using decision 1-3 was pruned to two nodes and decision tree using decision 1-5 was pruned to three nodes. The results are shown in figure 7a and figure 7b.

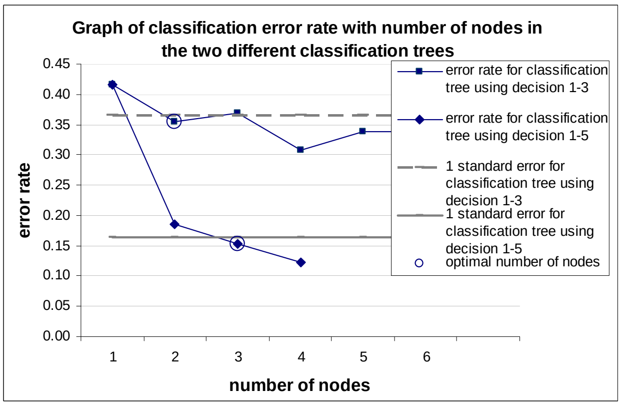
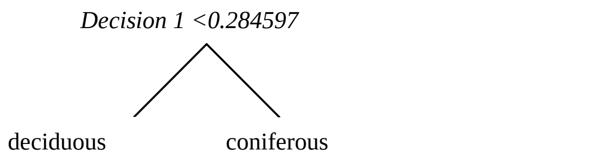


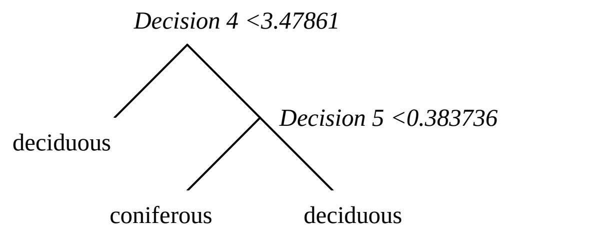
Figure 6 shows the decreasing error rate for the two classification trees with increasing number of nodes and the optimal number of nodes chosen for both trees displayed in figure 7a and figure 7b

Classification rate was 64.62% and 84.62% respectively. In addition, outliers are labelled and removed for each of the parameters in Figure 5. Seven trees were removed and the decision tree analyses were run again, the classification rates improved to 67.21% and 87.93% respectively.



Misclassification rate: 35.38%

Figure 7a shows the classification tree result by using decisions 1-3



Misclassification rate: 15.38%

Figure 7b shows the classification tree result by using decisions 1-5

3. 3D TREE MODELING

3.1 Internal structure reconstruction

In this section of the paper we address the reconstruction of the internal branching structures for individual trees. A similar approach was used for coniferous and deciduous bole derivation. However, for branch derivation, the number of classes for performing k-means clustering is defined differently for deciduous and coniferous trees. The purpose of this section is to use the results obtained from the previous chapter to derive bole and branches for the LiDAR tree crown. This approach is to try to get branching structures (leaf-off information) from leaf-on data, so to benefit from both types of data collection. When using LiDAR or aerial data to locate trees, one very common method is to estimate the location of the tree by the local maximum of the canopy height model (CHM) distribution (Dralle and Rudemo, 1996; Wulder et al., 2002). The assumption is that the boles of the trees grow straight from the ground; this might not be true for many trees. Therefore, tree trunks in this project are developed by using the entire point cloud profile.

3.2 Bole

Bole derivation is done by letting each point of the major trunk of the tree (assumed to have one major trunk per tree) is located in the geographic centre (x and y coordinate) of each predefined thickness (z) within the point cloud. This fixed thickness act like a moving window from the bottom to the top of the point cloud until no points can be read.

Then, the coordinates are connected to form the major trunk of the tree; Figure 8 shows the connected bole for two sample trees.

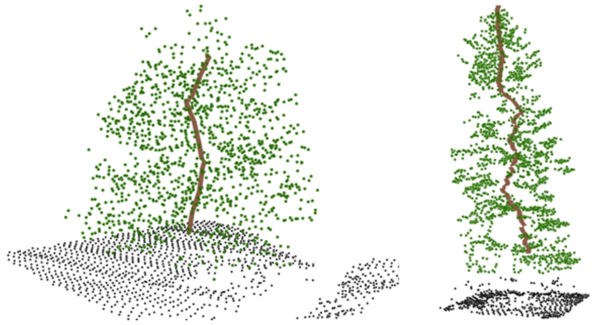


Figure 8. Figure showing the reconstructed bole for two sample trees

3.3 Branches

For deciduous trees, the branching structure derivation was done by k-means clustering using Euclidean distance in three dimensions. Before performing k-means clustering, the number of groups has to be determined; the separateness of the clusters can be calculated by finding the mean silhouette values for each k-means clustering. First, k means clustering was performed with number of groups equal to 2 – 50, then; the mean silhouette was calculated for each k-means clustering. The silhouette value was determined to find out the optimal number groups (number of k) for the set of point cloud. The smaller the silhouette value, the more separation there is among groups. For deciduous trees, the number of clusters is defined by the first local minimum for the mean silhouette. This is because the "general shape" of deciduous trees is hierarchical and to get the first order branch, the number of clusters should be large, but clusters should far apart enough to allow development of second order branches. The number of clusters which yield the first local minimum silhouette value should serve the purpose for deciduous trees because the number of clusters would not be large and the local minimums identify the relative importance of the separations among groups.

After the clustering was performed, the first order branches were defined as the closest (minimum Euclidean distance) point in each cluster joined to the furthest point in each cluster. This way, it formed the basic structure of the tree. After all the first order branches is derived, the entire first order structure is then transformed to the closest point in each cluster group and therefore creates the second order branches. Figure 9 shows a schematic diagram of the situation with two clusters.

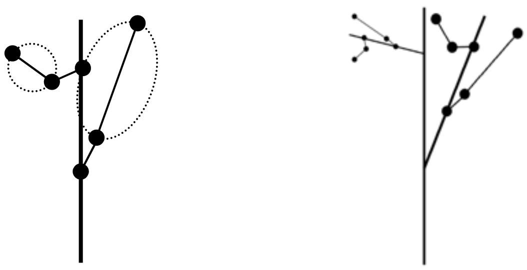


Figure 9. Schematic diagram of branch reconstruction for trees that are classified as deciduous, with number of cluster = 2, the left diagram connect the trunk point to the closest point and the furthest point in each

cluster and mimic the first order branch. The diagram on the right takes the left diagram and replaces the clusters mimicking the second order branches

For coniferous trees, branches extend out from the trunk forming multiple layers; thus, the number of clusters is expected to be large. Silhouette values decreases as the number of clusters increases; therefore, the numbers of clusters for coniferous trees are defined by the largest possible number of clusters that can be produced without any empty clusters within twenty iterations. After the clustering was performed, the branches are drawn by connecting by the furthest point of each cluster to the closest point on the trunk by Euclidean distance. Figure 10a shows the result of the clusters for a sample deciduous tree and figure 10b shows the result for branch reconstructions of the same tree. Figure 10c shows the result of the clusters for a sample coniferous tree and figure 10d shows the branch reconstruction for the same tree.

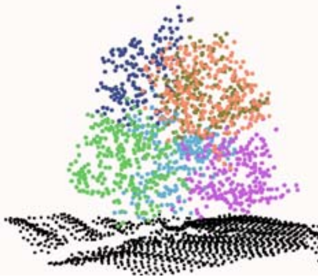



Figure 10a Figure 10b

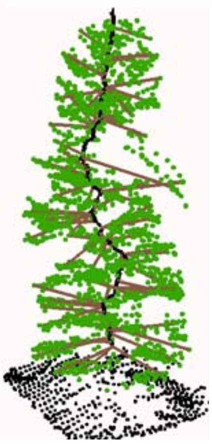



Figure 10c Figure 10d

Figure 10a shows different clusters in colours for a sample deciduous tree (in this example, number of clusters = 6). Figure 10b: Figure showing the location of derived bole and branches for the tree in figure 7a. Figure 10c shows the different clusters in colours for a sample coniferous tree (in this example, number of clusters = 80). Figure 10d shows the location of the bole and branches for the tree in figure 10c

4. DISCUSSION / CONCLUSION

There are two major types of LiDAR systems for research and commercial use, full waveform and discrete returns. This paper has used only the discrete returns of the range data, and by studying the geometry of the crown shape properties, we classified the different crowns into two major classes, deciduous and coniferous. From the classified results, bole and branching structures were reconstructed according to the type of tree. The shape of the tree crown is inherited in the gene (adaptation) and therefore a certain species will have the similar crown shape and branching structures. The other factor affecting crown shape and branching structure is the growing strategies, which is adopted by the growing neighbour environment and those are more difficult to model (Horn, 1971). As a result, crown geometry is believed to be an important piece of information for species classification.

By using just the three geometrical shapes (sphere, cone and cylinder), results were improved from 65% to 67% when the outliers were removed. If other parameters are included (area to volume ratio of convex hull and point to polygon buffering analysis), results were improved from 85% to 88%. Using crown shapes to classify trees is an intuitive method, but in this study it did not show promising results. By looking at the other geometrical properties, the results for classification increased considerably. Although different from what was expected, it is still believed crown shape and internal structure are good indicators for classifying trees, and future studies should be conducted in this direction.

This method of classification is quite simple to produced and arithmetically easy. Tree bole and branches reconstructions are for visualization, but can also be used to study growth behaviour and to provide insight regarding why trees grow in a particular directions. These results are useful in many types of studies. For example, it can be used to study the potential hazards of a tree growing into structures, by classifying trees into deciduous and coniferous provide a better growth estimates.

5. FUTURE DIRECTIONS

It is believed that tree crown geometry plays an important role in tree classification and should be continue to invest work on. In the future, it is expected to have better (more complex, more varieties) shape, more work will be done on developing internal structures of the L system tree to model real tree by integrating generative rules with data-driven information such as obtained from k-means clustering. Instead of having rigid rules resulting perfectly symmetrical tree crown, we would like to develop some flexible rules for tree crown detection. We also think that normalized distance for comparing LiDAR tree crowns and L system tree crowns is ineffective; having too much local variability and a different measurement method should be applied.

More work will also be done on convex hull analysis, including the complexity and orientation of the convex hull. On top of considering the different crown shapes for classification, like aerial photo interpretation, the roughness of the tree surface should also take into account. Optical and spectral data will be incorporated for verification and/or analyses. A sensitivity analysis will also be done to study individual tree parameters.

Full waveform sensors record the signals that represent the vertical structures of the targeted area as a time series, the time series are then modelled by different curves (e.g., Gaussian, log-normal) for physical property interpretations (Anderson et al, 2008; Koetz, 2006; Chauve, 2007). Riegl LMS-Q560 /LMS Q280i have full waveform recording capability which this project also has not taken into considerations but further research will. The trees selected for this paper have relatively complete tree crown point clouds (single isolated trees rather than a tree within a forested area), but in a forested area, the segmentation for individual trees for further clustering and classification can be problematic and should be taken into account in a subsequent study.

6. ACKNOWLEDGMENTS

The authors thank the Ontario Centres of Excellence (OCE), the Natural Sciences and Engineering Research Council of Canada (NSERC), and GeoDigital International Inc. for funding this research. Furthermore, we acknowledge Doug Parent for providing the valuable LiDAR data.

7. REFERENCE

Anderson, J. E., Plourde, L., Martin, M. E., Braswell, B. H., Smith, M.-L., Dubayah, R., Hofton, M. and Blair, J.B., 2008. Integrating waveform LIDAR with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, 112(4), pp. 1856-1870.

Brandtberg, T., Warner, T., Landenberger, R. and McGraw, J., 2003. Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density LiDAR data from the eastern deciduous forest in North America. *Remote Sensing of Environment,* 85(3), pp. 290-303.

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), pp. 325-340.

Breiman, L., Friedman J., Olshen R., and Stone C., 1984. *Classification and Regression Trees*. CRC Press, Boca Raton, FL. 372 p.

Chauve, A., Durrieu, S., Bretar, F., Pierrot-Deseilligny, M., Puech, W., 2007. Processing full waveform LiDAR data to extract forest parameters and digital terrain model: validation in an alpine coniferous forest, *Proceedings of the ForestSat Conference,* Montpellier, France.

Dralle, K. and Rudemo, M., 1996. Stem number estimation by kernel smoothing of aerial photos. *Canadian Journal of Forest Research*, 26(7), pp. 1228-1236.

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

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

Hopkinson, C., and Chasmer, L., 2007. Modelling canopy gap fraction from LiDAR intensity. *ISPRS Laser Scanning 2007 and Silvilaser*. Finland. IAPRS Vol. XXXVI, Part 3 / W52, pp. 190-195.

Horn, H. S., 1971. *The Adaptive Geometry of Trees*. Princeton University Press, Princeton, N.J. 144p.

Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X. and Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing,* 29(5), pp. 1339-1336.

Koetz, B., Morsdorf, F., Sun, G., Ranson, K. J., Itten, K. and Allgower, B., 2006**. I**nversion of a lidar waveform model for forest biophysical parameter estimation**.** *IEEE Geoscience and Remote Sensing Letters*, 3(1), pp. 49-53.

Liang, X., Hyyppä, J. and Matikainen, L., 2007. First-last pulse signatures of airborne laser scanning for tree species classification, Deciduous-coniferous tree classification using difference between first and last pulse laser signatures. *Proceedings of the ISPRS workshop, Laser Scanning 2007 and Silvilaser 2007*, Finland. IAPRS. Volume XXXVI, Part 3 / W52, pp. 253-257.

Lim, K., Treitz, P., Baldwin, K., Morrison, I., Green, J., 2003. LiDAR Remote Sensing of Biophysical Properties of Tolerant Northern Hardwood Forests. *Canadian Journal of Remote Sensing*, 29(5), pp. 658-678.

Litkey, P., Rönnholm, P., Lumme, J. and Liang, X., 2007. Waveform features for tree identification. In: *Proceedings of ISPRS Workshop, Laser Scanning 2007 and SilviLaser 2007*, Finland. IAPRS. Volume XXXVI, Part 3 / W52, pp. 258-263.

Persson, Å., Holmgren, J. and Sodermann, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*, 68(9), pp. 925-932.

Prusinkiewicz, P. and Lindenmayer, A., 1990. *The Algorithmic Beauty of Plants*. Springer-Verlag, New York. 228p.

Reitberger, J., Krzysteck, P. and Stilla, U., 2008. Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *International Journal of Remote Sensing,* 25(5), pp. 1407-1431.

Raber, G.T., Jensen, J.R., Schill, S.R. and Schuckman, L., 2002. Creation of digital terrain models using an adaptive LiDAR vegetation point removal process. *Photogrammetric Engineering and Remote Sensing*, 68(12), pp. 1307–1315.

Smith, W. K. and Brewer, C. A., 1994. The adaptive importance of shoot and crown architecture in conifer trees. *American Naturalist*, 143(3), pp. 528-532.

St-Onge, B., Vega, C., Fournier et Y. Hu, R. A., 2008. Mapping canopy height using a combination of digital stereo-photogrammetry and LiDAR. *International Journal of Remote Sensing*, 29(11), pp. 3343-3364.

Wulder, M., Niemann, K. O. and Goodenough, D., 2000. Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sensing of Environment*, 73(1), pp. 103-114.