Tree genera classification by ensemble classification of small-footprint airborne LiDAR

Connie Ko

A DISSERTATION SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN Earth and Space Science

YORK UNIVERSITY

TORONTO, ONTARIO

March 2014

Abstract

Tree genera information is useful in environmental applications such as forest management, forestry, urban planning, and the maintenance of utility transmission line infrastructure. The ability of small foot print airborne LiDAR (Light Detection and Ranging) to acquire 3D information provides a promising way of studying vertical forest structures. This provides an extra dimension of information compared to the traditional 2D remote sensing data. However, the techniques for processing this type of data are relatively recent and have becoming an innovative research direction. The existing perspective for processing LiDAR data for tree species classification involve calculating the statistics attributes of the vertical point profile for individual trees. This method however does not explicitly utilize the geometric information of the tree form such as shapes of the tree crown and geometric features that are derivable inside of the tree crown.

Therefore, the aim of this dissertation research is to derive geometric features from individual tree crowns and use these features for genera classification. The second goal of this research is to improve classification results by combining the newly developed features with the conventional vertical point profile features through ensemble classification system. Final goal of this research is to design a classification system to cope with the situation where the number of classes in the validation data exceeds the number of classes in the training data.

24 geometric features were initially derived and six of them are selected for the classification of pine, poplar and maple. Average classification accuracy of 88.3% is achieved by using this method. When the geometric features are combined with vertical profile features by ensemble classification system, the average classification accuracy increased to 91.2%. While the individual performance of geometric classifier and vertical classifier is 88.0% and 88.8% respectively for the classification of pine, poplar and maple. Lastly, when samples that do not belong to pine, poplar and maple are added to the validation data, the classification accuracy dropped to 72.8% by using randomly selected samples for training. However, through diversified sampling technique, the classification accuracy increased to 93.8%.

Acknowledgements

I would like to thank you to all of the persons who have supported and helped me in the past five years.

My sincerest gratitude is bestowed upon Dr. Gunho Sohn and Dr. Tarmo Remmel for their guidance, academic support, understanding, patience, and most importantly, their friendship during my doctoral studies at York University. I am extremely grateful to have them as my supervisors; it is such a great opportunity for me to do my doctoral program under their guidance and to learn from their expertise. I thank you them for encouraging me to grow into a better scientist, helped me through many challenges to my life.

I would like to thank you my examination committee members, Dr. Qiuming Cheng, Dr. Baoxin Hu and Dr. Richard Bello. Also, I thank you Dr. John Miller, Dr. Costas Armenakis and Dr. Aijun An for their contributions, suggestions, insights and encouragements to my research.

I would like to thank you all the GeoICT members for countless support throughout this research. I especially thank you Yoonseok Jwa, Brian Kim, Jili Li, Junjie Zhang and Nakhyn Song at Inha University, South Korea for acquiring field surveying data for this study for obtaining the field data collected in the summer of 2009 and 2011.

Very special thank you to all the faculty, staff and student members in the Department of Earth and Space Science and Engineering and the Department of

Geography. For all the support, encouragement, companionship and friendship you have given me.

This research was funded by GeoDigital International Inc., Ontario Centres for Excellence, and a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada. I would like to thank you Dr. Richard Pollock, Konstantin Lisitsyn, Doug Parent and Yulia Lazukova at GeoDigital International Inc. for their assistance in preparing the data. Also thank you CAG (Canadian Association of Geographers) for traveling funding to their annual meetings.

Lastly, I would like to thank you my husband, Brian Hung, my daughter Kaitlyn Hung and Kimberly Hung. I also thank you my parents, my sister for their patience, understanding, sacrifices and unconditional support throughout my academic life.

Table of Contents

List of Tables

List of Figures

List of acronyms and symbols

2D	two dimensional
3D	three dimensional
ALS	Airborne laser scanning
A(n)	Actual cumulative MDA, where n is the number of features removed
A_a	Area of the alpha shape of the tree crown
A_h	Area of the convex hull of the tree crown
A_{xy}	Area of tree crown projected to x-y plane
a_{xy}	Angle between each line segment to xy plane
$\vec{q}$	x and y component of $\vec{p}$
Bw	White birch
b_{xy}	Angle between the projected line to y-axis
c	Speed of light = $3 \times 10^8 ms^{-1}$
CV	Coefficient of variation
CVFDT	Concept-adapting very fast decision tree
δR	Change in residual distances
Δt_L	Time resolution
D	Dense forest structure
D_a	Diameter aperture of the receiver optics
d_h	Orthogonal distance from each point to the closest triangular facet of the convex hull
d_l	Orthogonal distance from each point to the best-fit line
d_p	Orthogonal distance from each point to the best-fit plane
DTM	Digital Terrain Model
ϵ	Error produced by base classifier
e	Estimated OOB error that is recorded for each classification tree
ECOC	Error correcting output codes
e_f	OOB error at f^{th} feature
F	First of many LiDAR return
FRI	Forest Resources Inventory
GIS	Geographic information system
GPS	Geographic positioning system
H	Flying height for the aircraft
H_c	Tree crown height
h_i	Base classifier that makes binary decision
h_p	Classifier that will produce class labels of pine and non-pine
h_o	Classifier that will produce class labels of poplar and non-poplar
h_m	Classifier that will produce class labels of maple and non-maple
H_t	Tree height
IMU	Inertial measurement unit

List of acronyms and symbols

$\vec{v}$	Normal vector of the closest triangular facet with a given LiDAR point
J	Objective function for SBS
k	Number of class, cluster
L	All possible predicted labels for a classifier
L	Last of many returns
La	Larch
l_i	Linear function through the 1^{st} to n^{th} features removed from SBS
LiDAR	Light Detection and Ranging
l_j	Linear function through the n^{th} to f^{th} features removed from SBS
L_n	Length of the line in the cluster
m	Class label maple
m'	Class label non-maple
M	Moderate forest structure
MDA	Mean decrease accuracy
MG(X)	Margin calculated based on X, where X is a vector contain selected features
MS	Multiple Strata
Mtry	Numer of feature variables randomly sampled at each split for Random Forests
η_{sys}	System transmission factor
η_{atm}	Atmosphere transmission factor
N_c	Number of points in the crown
N_{ij}	Count of points captured by i^{th} and j^{th} spheres
N_k	Number of points in cluster k
Nodesize	Minimum size of terminal node
Ntree	umber of trees generated within each iteration by Random Forests
o	Class label poplar
o'	Class label non-poplar
O	Open forest structure
OOB	Out of bag
OVA	One vs all
OVO	One vs one
$\vec{p}$	Direction vector of the best-fit line passes through cluster *k* pointing away from the center of the point cloud
PCA	Principal Component Analysis
PG(X)	Pseudo-margin calculated based on X, where X is a vector contain selected features
p	Class label pine
p'	Class label non-pine
P_i	Predicted *MDA* for the best fit line l_i

List of acronyms and symbols

Pj	Jack pine
P_j	Predicted *MDA* for the best fit line l_j
Po	Poplar
Pr	Red pine
P_r	Received power for LiDAR sensor
P_t	Emitted power for LiDAR sensor
Pw	White pine
R	Distance between the reflecting object and the sensor
R_n	New summation of residual distances after merging
R_o	Summation of residual distances prior to merging
ROW	Rights of way
S	Single return
SBS	Sequential Backward Selection
SD	Standard deviation
SEA	Streaming ensemble algorithm
SS	Single Stratum
S/N	Signal-to-noise-ratio
SVM	Support Vector Machine
Sw	White spruce
T	Number of classifiers
t_L	Travelling time of a light pulse
t_{rise}	Rise time for pulse ranging
UFFT	Ultrafast forest tree
V(X)	Average votes calculated based on X, where X is a vector contain selected features
V_a	Volume of the alpha shape of the tree crown
V_h	Volume of the convex hull of the tree crown
v_g	Group velocity of the laser pulse
V_{ij}	Overlapped volume between i^{th} and j^{th} speheres
V_k	Convex hull volume for cluster k
$\vec{w}$	Vector from the LiDAR point to one of the closest convex hull vertices
WCE	Weighted classifier ensemble
X	Features selected for classification, where $X \subset \mathbb{R}^f$
Y	Correct class label obtained from field validation
y	Predicted class label
y*	Final predicted class label (y*) from ensemble classification
σ	MG boundary in the validation data that separates trees that are potentially incorrect from the trees that are potentially correct
$\sigma(R)dR$	Apparent effective differential cross-section
λ	Wavelength

Chapter 1 Introduction

1.1 Research Motivation and Problem Domain

The measurement of forest attributes such as tree height, stem density, is important in monitoring the forests. Most importantly, species or genera information is valuable for the management of forest in predicting production yield and growth and to describe forest ecosystems. The species / genera information not only is important in updating the forest resource inventory (FRI), knowing the genera can also have better estimates on tree growth rate and is important in the management of vegetation around human infrastructure. One example is the vegetation management of the Right of Ways (ROWs) near power transmission lines, such that the management companies can have better estimates on trimming or clear cutting schedules. The measurements of attributes can be obtained by field surveying or remote sensing techniques. Field surveying provide a direct measurement of the tree attribute but is extremely time and cost ineffective, as a result, remote sensing technology can be beneficial tools for obtaining this information.

Tree species / genera information can be obtained (or inferred) from spectral signatures of remotely sensed imagery, especially when the hyperspectral information is given (Clark et al. 2005; Cochrane, 2000; Buddenbaum et al. 2005). In order to yield detailed map, higher spatial resolution aerial photographs can be obtained allowing species classification for individual tree crown (Brandtberg, 2002). These imagery based methods extract tree species information from two dimension data and have been proven

effective. In fact, the standard means of collecting highly accurate forest inventory data still relies on ground surveys (cruising) or photogrammetric methods (Gillis and Leckie, 1996). In Ontario, the Forest Resources Inventory (FRI) is created by the Ministry of Natural Resources. FRI is standardized and stored in a geographic information system (GIS) that is updated from aerial photography (Leckie and Gillis, 1995). Developments in geoinformatics technologies (e.g., GIS data, remotely sensed data) are providing additional opportunities for detailed updates to the FRI (e.g., Gillis, 2001; Morrison et al., 1999; Hilker et al., 2008; Wulder and Seemann, 2003). Until the relatively recent development of LiDAR (Light Detection and Ranging) data provided new insights in remote sensing as it provide three dimensional data. In Woods et al. (2011), the authors discuss the applications and issues of large scale Light Detection and Ranging (LiDAR) operational in Ontario boreal forests in terms of lowering the cost on an "area" based classification. Indicating that for individual tree classification, a higher point density data is required.

The motivation of this dissertation is to investigate the possibility of using high density LiDAR for individual tree genera classification. As LiDAR technology matures, the point density acquired for the recent studies are becoming higher. The existing methodologies of tree classification using LiDAR can be categorized into three groups and will be discussed further in Chapter 2. The first group is a top centric approach including researches that combines spectral information with LiDAR data. The second group of research studies the vertical point profile of the tree and the third group of research utilizes the geometric information of LiDAR point distribution reflected from

the individual tree. Geometric information is especially useful in representing the form of a tree and can be easily related to the biophysical condition. The top-centric approach was not able to highlight the LIDAR signals that have reached the lower part of the tree; the vertical point profile approach is computationally efficient but does not tie close relationship to the biophysical implications of tree form as compared to the geometric approach. This research focuses on the geometric approach and will compensate the existing geometric approaches where geometric features relate to the outer shape of the tree crown (e.g. curve fitting, alpha shapes). LiDAR points reflected from the lower and inner part of the tree crown is as important as the points reflected from the surface of the tree crown. However, only limited amount of research has been performed in this direction with airborne systems, most of them are based on terrestrial systems (Busksh and Fleck, 2011; Pfeifer and Winterhalder, 2004; Park et al., 2010; Xu et al., 2007; Côté et al., 2009, 2011) where point density per tree is significant. The biggest drawback on using terrestrial systems on forest applications is that it only covers a small area whereas aerial systems can cover a larger area with a lower cost. As a result, this research explore the potential of using airborne LiDAR data for developing geometric features that are related to the internal (branching structures) and external (overall shape of the tree) tree crown shape (Chapter 3).

Ensemble classification is an emerging strategy that is being used for improving classification accuracy. The advantage of using ensemble classification is that the training data can be trained independently by the individual base classifiers. The base classifiers learn about the classification independently and therefore potentially makes the same or

different decisions. By combining the decisions wisely, classification accuracy can be improved (Chapter 4). The multiple classifier systems (ensemble classifiers) are especially appropriate for the inventory of large scale data because individual classifier can be trained on independent dataset where each one represents the different geographic locations. Trees grow differently in different environmental conditions even though the overall tree form may appear similar. The benefit of selecting a wide variety of training data from different geographic locations is that the base classifiers can learn with maximum diversity. The benefits and discussions on training data diversity are included in Chapter 5.

1.2 Applications of LiDAR

Airborne LiDAR have many field of applications; some research use bathymetric LiDAR for coastal boundary mapping (Lefsky et al., 2002 and Irish et al., 1999) and some for Digital Terrain Model (DTM) generation (such as Kraus and Pfeifer, 2001 and Brovelli et al., 2003). LiDAR is also popular for urban studies because of the rich three dimension information provided, allowing many research in building reconstructions (for example: Becker and Haala, 2007; Malllet et al., 2008; Schenk and Cshato, 2007; Baillard and Maître, 1999 and Rottensteiner et al., 2005). Other than the above applications, LiDAR has also been used widely in forest applications particularly in retrieving tree height information as tree height is one of the most important biophysical parameters in forest studies. The history of airborne LiDAR on forest study can trace back to 1980s and even further, Nelson et al. (1984) suggested using laser to derive height profile for retrieving forest characteristics. Aldred and Bonnor (1985) suggested

the use of waveform for canopy height estimation and since then, a lot of research has been developed and conducted aiming at the acquisition of tree height, canopy volume, biomass estimation, growth estimation and many more mainly for forest management. These researches have two main approaches, a distribution based approach and individual tree based approach (Hyyppä et al., 2008). Distribution based approach consider forests as points of distribution, therefore, the forest attributes (such as height, biomass volume) are obtained using statistical method, examples are: Næsset (2004); Hollaus (2006) and Means et al. (1999). These methods could be area sensitive and therefore calibration is needed. However, these approaches are especially efficient if the study area is large. Another approach is individual based approach, where the forest attributes are collected or calculated on a single tree basis, examples are: Reitberger et al. (2007); Solberg et al. (2006) and Korpela et al. (2007). These studies can be arithmetically expensive due to complexity in the algorithms and the large amount of data points but the results are much more in detailed, the objective of this dissertation is in-line with this type of research.

1.3 Research Objectives

The three main objectives of this research are discussed in three chapters.

1. Chapter 3: The first objective of this dissertation is to develop methodologies for using LiDAR data for single tree genera classification. Geometric features were derived as attributes for classification. In contrast to the conventional method using vertical profile attributes, geometric features provide a better understanding, visualization and classification results. The chapter discusses both internal and

external geometric features, where internal features tackle features inside of the tree crown such as locating the branches. On the other hand, external features tackle features related to the boundary of the tree crown such as the overall shape of the tree. The methods developed can be applied to vegetated areas such as forests.

2. Chapter 4: In this chapter, two classifiers are to be built from two sets of features, the first is from geometric features (Chapter 3) and the second is from the vertical profile features. The second objective involves two tasks; the first is to reduce the number of features for the two classifiers. This is to reduce the model complexity by selecting only the important features in each classifier. The second task is to combine the two classifiers by ensemble techniques in order to improve classification accuracy. Ensemble classification a classification strategy that combines the decisions from multiple classifiers for making the final decision, this approach combines the advantage of both classifiers.

3. Chapter 5: The third objective of this research is to study the effects of inconsistent quality found in the collected LiDAR tree samples. The inconsistency could be a result of occlusion or tree crowns overlapping each other where single tree crowns are unable to be isolated clearly. This chapter first quantify sample quality and study the distribution of quality over the training sample. By experimenting with different training sample distributions, the classification accuracy can be improved by including the most diversified sample quality.

Additionally, this chapter improves the practicality of the classification scheme by considering the existence of “unknown” instances. A common problem in the classification of environmental objects is that the number of classes in the validation sample exceeds the number of classes in the training sample. Ideally, an “unknown” class will be generated in the classification results categorizing all the data that does not belong to any of the training classes. However, this is not an easy task because of two reasons. First, typically the “unknown” class will not be available for training. Therefore classifier is not able to learn about the feature distribution of the “unknown” class. Second, the “unknown” class normally include a variety of objects, meaning a variety of feature distributions, that are difficult to group properly. This goal is achieved by decomposing the multi-class classification problem into a series of binary classification problems.

In summary, the thesis of this research is that the individual tree crowns isolated from LiDAR point cloud can be automatically classified into three genera, four categories (pine, poplar, maple and “unknown”). Four classification labels will be generated including an “unknown” label that includes tree that does not belong to the genera provided by the training data.

1.4 Methodology Overview

Section 1.3 describes the research objectives and this section provides the methodology overview to achieve the above objectives.

1. In Chapter 3, few experiments were performed by the following methods to achieve objective 1:

    a. 24 geometric features are derived in Chapter 3. First, points recorded for each tree crown are clustered into representative groups by a merging-and splitting algorithm. They are grouped according to their spatial organization of the points and each group is designed to represent one main branch of the tree. Then, a best fit line and best fit plane is fitted through each cluster. For volumetric features, alpha shapes and convex hulls are also calculated for each cluster and for the entire tree crown. The orthogonal distances from each point to the closest facet of the tree crown convex hull is also calculated. Next, each point is buffered outward to become an individual sphere, the overlapped volume and count is recorded for each tree crown. Lastly, ratios such as tree crown height divided by tree height for the overall tree form are calculated.

    b. A sensitivity analysis is performed to determine the optimal training size needed for the classification. Data is partitioned randomly from 5% to 95% for training and the classification accuracy is recorded for each partition.

    c. A geometric classifier is built based on the 6 most significant geometric features using Random Forests algorithm. The mean decrease accuracy is

calculated for each feature for assessing the contribution of each feature to the classification.

d. Classification accuracy is calculated for individual field sites

2. In Chapter 4, the following experiments were performed by the following methods to achieve objective 2:

   a. 78 Vertical point profile features were calculated. These features summarize the properties of return types, mean, standard deviation, kurtosis of variation, skewness of variation for height and intensity values. These calculations are performed for at the 10^{th} height percentile, 50^{th} height percentile, 90^{th} height percentile and for the entire canopy.

   b. 24 geometric features and 78 vertical profile features are reduced to 6 and 26 respectively for simplifying the model complexity. This is performed in two steps, first, features with high correlations are removed. Next and objective function is set up for Sequential Backward Selection method, using the mean decrease accuracy values obtained from Random Forests. The optimal number of features for each set of features is selected by minimizing the pre-defined objective function *J*.

   c. An ensemble classification system is built for combining two sets of features (geometric features derived by Chapter 3 and vertical profile features calculated in this chapter). A sequential followed by a parallel

model is built, first validation samples are judged by geometric classifier. Potentially mis-classified trees are filtered out automatically. These trees are assessed again by both geometric and vertical profile classifier and the final decision for these samples is made based on the comparisons of the margin differences between two classifiers. The classification accuracy is calculated for 1) geometric classifier alone 2) vertical profile classifier alone and 3) ensemble classifier for the purpose of comparison.

d. A sensitivity analysis is performed to study the relationship between classification accuracy and LiDAR point densities. This is also done to investigate the lower limits of point density for this classification scheme, potentially lowering the cost of data acquisition in the future.

3. In Chapter 5, the following experiments were performed by the following methods to achieve objective 3:

   a. 3 binary one-versus-all classifiers are created for the classification of pine versus non-pine; poplar versus non-poplar and maple versus non-maple. A parallel ensemble classification system is built for combining the three binary classifiers together. This method is to achieve the objective of classifying “unknown” classes in the validation data.

   b. The quality of each LiDAR tree is quantified by the ratio between the number of times a sample is being classified correctly and the number of times a sample is being selected as out-of-bag data from Random Forests.

A sensitivity analysis is performed to study the relationship between the classification accuracy and the distribution training data quality. Result shows training data contained the most diversified quality yield the best accuracy rate.

c. Classification is performed using 1) random sampling method with ensemble classification 2) diversified sampling method with ensemble classification and 3) diversified sampling method with Random Forests classification.

4. Chapter 6 concludes and summarizes this research. Figure 1-1 shows the structure and organization of the dissertation.

1.5 List of Publications

Below shows a list of publications associated with this research:

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

Ko, C., T.K. Remmel, G. Sohn and J.R. Miller. 2013. Ensemble classification of tree genera from airborne LiDAR point clouds. Silvilaser, Oct 9 – Oct 11, Beijing, China.

Ko, C., T.K. Remmel, and G. Sohn. 2012. Mapping tree genera using discrete LiDAR and geometric tree metrics. BOSQUE 33(3):313-319.

Ko, C., G. Sohn, and T.K. Remmel. 2012. The impact of LiDAR point density on classifying tree genus: using geometric features and vertical profile features. SilviLaser 2012, September 16-19, Vancouver, British Columbia, Canada.

Ko, C., T.K. Remmel, and G. Sohn. 2012. Algorithm for processing a LiDAR point cloud to retrieve internal geometric tree crown structures. Canadian Association of Geographers Annual General Meeting, May 28 - June 2, Waterloo, Ontario, Canada.

Ko, C. and T.K. Remmel. 2012. Forest genera mapping using discrete LiDAR and geometric tree metrics. IUFRO Landscape Ecology Conference, November 5-12, Concepción, Chile.

Ko, C., G. Sohn, and T.K. Remmel. 2012. A comparative study using geometric and vertical profile features derived from airborne LiDAR for classifying tree genera. The XXII Congress of the International Society of Photogrammetry and Remote Sensing, 25 August - September 1, Melbourne, Australia.

Ko, C., T.K. Remmel, and G. Sohn. 2011. Extracting geometric features from airborne LiDAR for tree species classification. Canadian Association of Geographers Annual General Meeting, May 31 - June 4, Calgary, Alberta, Canada.

Ko. C, T.K. Remmel and G. Sohn, 2010. A statistical partitioning of vegetated airborne laser scanning data towards understory and canopy separation. Canadian Association of Geographers Annual General Meeting, June 1-5, Regina, Saskatchewan, Canada. (Proceedings)

Ko. C, G. Sohn and T.K. Remmel, 2010. Derivation of 3D geometric models of single tree using high density airborne LiDAR data. AOLS (Association of Ontario Land Surveyors) 118th Annual General Meeting, Feb 17-19, 2010, Huntsville, Ontario.

Ko. C, G. Sohn, T. Remmel, 2010. Experimental investigation of geometric features extracted from airborne LiDAR for tree species classification. Silvilaser 2010 – 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems, September 14 to 17, 2010 – Freiburg, Germany.

Ko. C, G. Sohn and T. Remmel, 2009. Classification for Deciduous and Coniferous Trees Using Airborne LiDAR and Internal Structure. Silvilaser 2009, Oct 14-16, 2009, Texas A&M, USA.

Ko. C, G. Sohn and T.K. Remmel, 2009. A Deciduous-Coniferous Classification and Internal Structure Derivation Using Airborne LiDAR Data. ISPRS Laserscanning 2009, September 1-4, Paris.

Ko, C., Remmel, T.K., and G. Sohn. 2009. A precise reconstruction of three-dimensional crown shape from airborne LiDAR point clouds. Canadian Association of Geographers Annual General Meeting, May 26-30, Ottawa, Ontario, Canada.

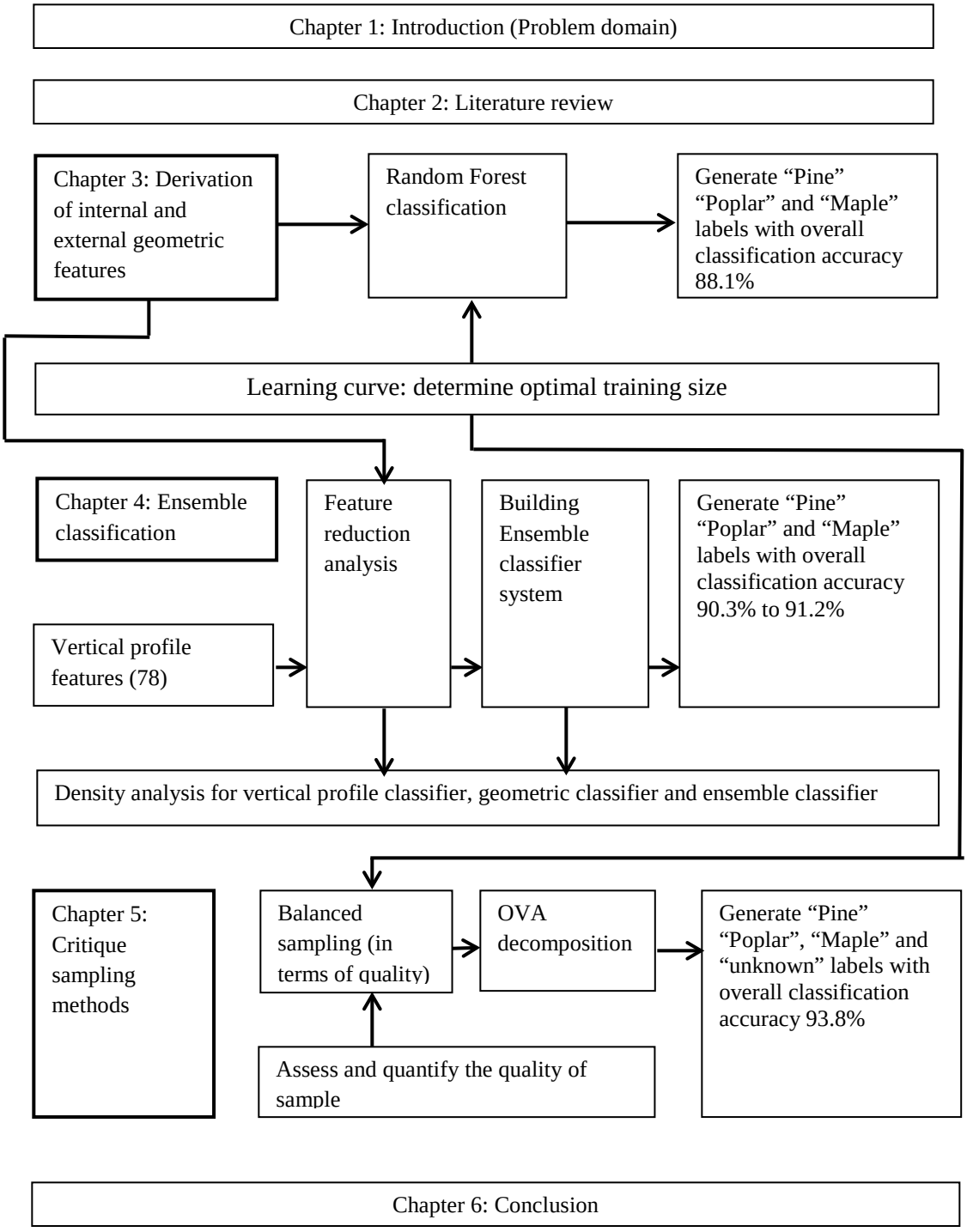

Figure 1-1 Structure of the dissertation

Chapter 2 Background

2.1 Introducing the research – present knowledge

2.1.1 Light Detection and Ranging (LiDAR)

Light Detection and Ranging (LiDAR) is an active remote sensing sensor based on distance (range) measurements and can be mounted on an airborne device for data acquisition. The detection systems are categorize into waveform digitization (or continuous-wave) and discrete return. Waveform digitization produces a continuous waveform (a time series) that fully describes the vertical structure of the ground targets whereas discrete return detected peaks when the return signal exceeds a detection threshold. Therefore discrete return system produces discrete points throughout the vertical profile of the structure (usually up to five multiple returns per pulse). The data type used for this dissertation belongs to discrete return system; as a result, the following discussion will focus on the discrete return system.

For discrete return system, or pulse ranging, range is determined by the time of pulses emitted from the sensor and the return of the reflected signals. On the other hand, continuous wave systems measures range by the phase difference between the transmitted and received signal (Mallet and Bretar, 2009). Equation (2.1) to (2.4) summarizes the fundamental parameters that describe the discrete systems (Wehr and Lohr, 1999).

$$Range = R = \frac{1}{2} \times c \times t_L \tag{2.1}$$

$$Range\ Resolution = \ \Delta R = \frac{1}{2} \times c \times \Delta t_L \quad (2.2)$$

$$Maximum\ Range = \ R_{max} = \frac{1}{2} \times c \times \Delta t_{Lmax} \quad (2.3)$$

$$Range\ Accuracy = \ \sigma_n = \frac{c}{2} \times t_{rise} \times \frac{1}{\sqrt{S/N}} \quad (2.4)$$

Where t_L = travelling time of a light pulse

R= distance between the reflecting object and the sensor

c= speed of light = $3 \times 10^8 ms^{-1}$

Δt_L = time resolution

t_{rise} = rise time for the pulse ranging

S/N = signal-to-noise-ratio

From equation (2.2), it shows that the range resolution ΔR is directly proportional to the time resolutionΔt_L, however, for airborne laser scanning (ALS), the resolution is less important because in most cases it is much lower than the range accuracy (Baltsavias, 1999). For discrete return systems, the maximum unambiguous range depends on two factors. The first is the maximum range (the number of bits) of the time interval counter and the pulse rate where the maximum time interval can be measured by the time counter of the laser (Baltsavias, 1999). In practice, to avoid the confusion between pulses, a pulse has to be received before another pulse can be sent. Normally, the time interval is large enough so that the maximum range is affected by laser energy losses rather than Δt_{Lmax}. Therefore is affected by factors such as laser power and beam divergence, atmospheric transmission, target reflectivity, etc. (Wehr and Lohr, 1999;

Baltsavias, 1999). Range accuracy σ_n is affected by the rise time for the pulse and inversely proportional to the signal-to-noise ratio (S/N), where S/N is affected by the power of the received signal, input bandwidth, background radiation and other factors controlled by the systems. (Wehr and Lohr, 1999). According to Mallet and Bretar (2009) and Wagner et al. (2006), the reflected signals for targets are the superposition of echoes and therefore can be expressed as equation (2.5).

$$Received\ power = P_r(t) = \frac{D_a}{4\pi\lambda^2}\int_0^H \frac{\eta_{sys}\eta_{atm}}{R^4} P_t\left(t - \frac{2R}{v_g}\right)\sigma(R)dR \qquad (2.5)$$

Where t = time

D_a = diameter aperture of the receiver optics

P_t = emitted power

λ = wavelength

H = flying height for the aircraft

R = distance between the reflecting object and the sensor

η_{sys} = system transmission factor

η_{atm} = atmosphere transmission factor

v_g = group velocity of the laser pulse

$\sigma(R)dR$ = apparent effective differential cross-section

It is also worth to note that although this research relies on data collected airborne, LiDAR data can be collected as ground based systems as well, called terrestrial LiDAR. Terrestrial LiDAR operates under the same principle described in equation (2.5)

and is suitable for studies that have a smaller area of interest and data are collected laterally whereas airborne LiDAR is suitable for studies that have a larger area of interested and are collected aerially. For airborne LiDAR, the location of the flight path is usually documented by differential GPSs (GPS stations on the ground) and the precise measurement of the sensor location and orientation (mounted on the aircraft) is recorded by the inertial measurement unit (IMU), where the yaw, pitch and row attributes are recorded. The sensor sends out laser signals and measure the time needed for a laser signal to return back from ground target; knowing the speed of light, range can be calculated for each signals. Common wavelengths for measuring vegetation is usually 900nm, 1064nm, 1470nm and 1560nm due to the high reflectance in those regions and the pulse width is usually 7ns wide (full-width at half the maximum amplitude) (Petrie and Toth, 2009).

2.1.2 Tree species/genera classification using airborne LiDAR

The use of LiDAR for forestry applications has become popular in the past decades with LiDAR technology becoming more available. Existing and related studies can be categorized into three major groups with different perspectives of looking at the LiDAR data in relation to vegetation. The first group focuses on extracting information near or at the tree top. This tree top centric approach assumes that the most important characteristics for crown classification are contained in the upper and central portion of the crown. Holmgren and Persson (2004) found that point density generally decreases as depth from the canopy top increases; therefore, more information for classification is available at the upper part of the crown. Rahman and Gort (2008) and Rahman et al.

(2009) reported that the point density also decreases from the center of the tree outward, meaning that most of the information is contained in the central portion of the crown. Thus, some studies retrieved the information for classification features from the surface of the tree crown by using the combination of LiDAR data and spectral information (Persson et al., 2004; Koukoulas and Blackburn, 2005; Hill and Thomson, 2005; Hilker et al., 2008; Holmgren et al., 2008). These approaches have demonstrated successful and promising perspectives in species classification. Recently, LiDAR sensor technology has rapidly advanced, which achieves a higher foliage penetration rate and provides more opportunities to retrieve the information from inside of crown. In this regard, the tree top centric classification methods cannot highlight the LiDAR signals that have reached the lower part of the tree, and it might bias points at or near the top of the forest canopy, overlooking the importance of points within crowns.

To overcome the problem of biasing the tops of crowns and to emphasize the importance of LiDAR points reflected from the lower part of the canopy, the second group of research focuses on the retrieval of tree species information and has considered internal points for the analysis of the vertical height and (or) intensity distributions of LiDAR point clouds (Holmgren and Persson, 2004; Brandtberg, 2007; Ørka et al., 2009; Kim et al., 2009; Korpela et al., 2009; Suratno et al., 2009). These studies measured the frequencies, means, and standard deviations of first and single returns for the entire tree height profile (Brandtberg, 2007; Kim et al., 2009) or by height percentiles (Holmgren and Persson, 2004; Ørka et al., 2009; Korpela et al., 2009). These vertical profile methods show that it is important to address the LiDAR points that were reflected from the lower

part of the canopy or from within the crown. Studies that utilize the entire vertical profile take advantage of LiDAR's ability to penetrate tree canopies, permitting the calculation of attributes related more closely to the entire tree, whereas characteristics related to the percentiles specify the importance of point distributions relative to ranges of tree height. Both perspectives support the use of LiDAR points located deep in the crown. Moreover, with better equipment, allowing a lower flying altitude and stronger LiDAR signals, a dataset with higher point density can be obtained. This provides a better method of studying the relationship between points reflected from within the crown and genera information, as more points are recorded from lower in the crown. A limitation of using the vertical profile method is that the information retrieved from point distributions does not explicitly consider the geometric characteristics of tree crown. Tree form, however, is important and should be taken into consideration to ensure accurate tree classification.

The third group of research focuses on investigating and extracting geometric information of the outer surface of the tree crown from point clouds. Barilotti et al. (2009) fit curved surfaces to tree tops of different species to consider canopy geometry, and Kato et al. (2009) used an iso-surface algorithm to wrap surfaces around crowns for retrieving parameters like crown volume, width, and base height. Considering tree crowns as individual objects, some studies have resorted to computing representative alpha shapes for individual trees, with boundaries that enclose a set of points similar to a convex hull but that allow internal curvature (Vauhkonen et al., 2008; Vauhkonen et al., 2009; Vauhkonen et al., 2010). These approaches are an improvement over simple segmentation of points with height percentiles, and suggest that geometric information is

important for tree species classification. The features derived from convex hulls, alpha shapes, and parabolic fitting focus on the overall shape of trees (external features) with LiDAR pulse density ranges from 4 returns per m^2 (Barilotti et al., 2009) to 40 returns per m^2 (Vauhkonen et al., 2008, 2009). The aforementioned studies have concentrated on geometric characteristics of the outer surface of tree crowns. However, using a hand-held camera, Niccolai et al. (2010) showed that information about the colour, shape, patterns of tree form (stems and branches), and leaves within a crown are considered critical components at the crown level.

This study used a high pulse density of 40 pulses per m^2 dataset, with up to five returns per pulse, putting it in line with the high density studies; this allows us to consider geometry derived from within the crown and to use this information for genera classification. Chapter 3 of this dissertation will focus on the derivation of geometric features for genera classification.

2.1.3 Ensemble classification systems

Ensemble methods in classification is the training of multiple classifiers to solve the same problem, in different community, it is also called committee-based learning, multiple classifier systems, mixture of experts etc (Samadzadegan et al. 2010; Zhou, 2012; Ruta and Gabrys, 2000). The criteria of a good ensemble system should provide an increase in classification accuracy. There are numerous ways of combing classifiers suggested in the previous research (Oza et al. 2005), with a wide variety of applications such as text categorization (Bell et al. 2005), speech recognition (Linares et al. 2003) and

hand written words recognition (Koerich et al. 2005). In particular, a few example of such studies in remote sensing includes Samadzadegan et al. (2010); Lodha et al. (2006); Kumar et al. (2002) and Kavzoglu and Colkesen. (2013).

Unlike the traditional classification learning where one learner train and classify the dataset, ensemble classification assemble various learners through different models and make decision in collaboration of all learners (Zhou 2012). In general, there are two ensemble models. The first is parallel ensemble model where base learners make the decision in parallel (an example will be bagging methods) and sequential ensemble model where base learners make decision in sequence (an example will be boosting methods). One of the objectives for combining classifiers is to reduce the overall error and hence increase overall classification accuracy as compared to a single classifier (Ali and Pazzani 1996; Breiman 1998; Dietterich 2000; Bryll et al. 2003 and Kavzoglu and Colkesen 2013).

Generally, parallel models decrease errors by combining base classifiers that are as independent as possible (diverse). Classifiers that have different perspectives can correct the errors made by the other base classifiers. For sequential models, errors are reduced in a residual decreasing way, if the subsequent base classifiers are able to correct the error made by the previous base classifiers, overall error will decrease (Zhou 2012).

In parallel models, Zhao (2012) has showed that errors reduce exponentially to ensemble size T by Hoeffding inequality using the following example:

Let $h_i$be the binary base classifier {+1,-1} with error ϵ such that

$$P\big(h_i(x) \neq f(x)\big) = \epsilon \tag{2.6}$$

And by combining *T* classifiers, one can form the ensemble classifier $H(x)$, and can be written as:

$$H(x) = sign(\textstyle\sum_{i=1}^{T} h_i(x)) \tag{2.7}$$

Assume the final decision is made by majority scheme, $H(x)$ will misclassify if at least half of the base classifiers make an error, therefore, the generalization error can be written as:

$$P\big(H(x) \neq f(x)\big) = \textstyle\sum_{k=0}^{T/2} \binom{T}{k} (1-\epsilon)^k \epsilon^{T-k} \leq \exp(-\frac{1}{2} T(2\epsilon - 1)^2) \tag{2.8}$$

This shows that the generalization error reduces exponentially as T becomes larger, also, as $T \to \infty, P\big(H(x) \neq f(x)\big) \to 0$.

On the other hand, in sequential models, the overall errors are reduced in a residual decreasing manner. An efficient sequential model has base classifiers ordered in a way that the subsequent base classifier is able to correct the mistake of the prior base classifier. Also, base classifiers will have to be diverse enough so that same mistake will not be made in the succeeding level. Ensemble methods are applied in both Chapter 4 and

Chapter 5. In Chapter 4, a hybrid of parallel and sequential model is built using Random Forests as base classifier. Chapter 5 decomposes a multiclass problem into a series of binary classification problems, ensemble methods are applied by combing the binary classifiers in parallel. The binary classifiers are also running Random Forests. The following subsections will review the methods of 1) decomposing a multi-class problem into binary classification problems. 2) Random Forests as an ensemble classifier.

2.1.3.1 Multi-class to binary classification

In Chapter 5, the multi-class classification is being broke down into a series of binary classification. There are two major motivations on ensembling the binary base classifiers. The first is the production of a class label "unknown" that categorizes all the validation data that does not belong to any class exhibited in the training data. The second is the flexibility of concept change, the removal or addition of extra class label can be done without changing the rest of the model, for future studies.

There are three poplar techniques for reducing a multi-class problem into a series of binary classification problems (Galar et al. 2011), the first one is called "one-vs-one"(OVO) (Hastie and Tibshirani 1998) and the second one is called the "one-vs-all" (OVA) (Rifkin and Klautau, 2004) and the third one is called the Error Correcting Output Codes (ECOC). OVO consists of binary classifiers that differentiate between two classes (for example pine vs maple) and therefore for a k class problem, $\binom{k}{2}$ classifiers are needed for ensemble. On the other hand OVA consist of binary classifiers that discriminate between particular classes from the rest of the classes (for example, pine vs

non-pine) and therefore k classifiers are needed for solving a k class problem. ECOC is a method where a code word is generated from a series of binary classifiers for each class label. For any validation sample, a code word from that sample will be generated and will be compared to the code word generated from the training samples. The class with the minimum Hamming distance will be labeled (Dietterich and Bakiri, 1995).

To compare the performance between OVO and OVA, study such as Galar et al. 2011 shows that although there are no significant difference found in OVA and OVO, but both strategies outperformed the original classifier. In another study, Hashemi et al. 2009 show that OVA attains better classification when compared with concept-adapting very fast decision tree (CVFDT), a single multiclassifier; weighted classifier ensemble (WCE), and streaming ensemble algorithm (SEA), both are ensemble of multiclass classifiers; and ultrafast forest tree (UFFT), an OVO method. In Hsu and Lin (2002), they show that OVO outperformed OVA, however, in Rifkin and Klautau (2004) study, authors suggested that OVA is performing just as well as OVO. These literatures show that neither OVO nor OVA consistently outperform one another. Instead, these studies indicate that the decomposition of a multiclass problem into series of binary classification problems is an efficient approach and often outperformed the original multi-class classifier. In fact, OVO and OVA are poplar methods for combining Support Vector Machine (SVM) classifiers (Duan et al. 2007; Milgram et al. 2006; Hsu and Lin 2002; Liu and Zheng 2005)

2.1.3.2 Random Forests classification

Random Forests is an effective classification algorithm and is being applied in various experiments in the dissertation. Random Forests is an ensemble classifier comprised of many classification trees for categorical data, or regression trees for continuous data trained from a subset of the data (Breiman, 2001, Liaw and Wineder, 2002). Unlike classification tree or regression tree analysis, where a single classification tree is produced for classification, Random Forests constructs a user-defined number of trees, *Ntree*, by using randomly selected feature variables.

For each iteration of classification tree construction, a subset of the data is drawn by sampling with replacement (a bootstrap sample), where about 37% of the data (out-of-bag data or OOB data) is set aside for testing the classification accuracy of the current tree and 63% of the data (in-bag data) is used to construct a single decision tree. According to Zhou (2012), given *m* training examples, the probability of the i^{th} sample will be selected at least once can be approximate by Poisson distribution with λ=1 and therefore is $1 - 1/e \approx 63\%$. Therefore, about 37% of the original data have never been used for training. In Random Forests, the decision tree grows by consequently splitting a parent node into binary descendent nodes, by using the random subset of features. The best split at each node is defined by an optimal feature subset that minimizes the class impurity within the descendant nodes.

Random Forests classification has successfully been employed in forestry applications. For example, Falkowski et al., (2009) used 34 features derived from LiDAR, then reduced the number to 9 to classify successional forest stages. Martinuzzi et

al., (2009) used Random Forests to classify snag types by deriving 34 LiDAR features (19 of them related to the canopy height point distribution and 15 of them related to the topographic information). The implementation for this dissertation is performed within the *randomForest* package for R (R Development Core Team, 2013; Breiman, 2001).

The main input parameters of Random Forests for classification include:

1. Training sample labeled with known species and geometric features
2. the number of feature variables randomly sampled at each split (*mtry*, approximately $\sqrt{number\ of\ features}$) (Strobl et al., 2009)
3. the number of trees generated within each iteration (*Ntree*, 1000 for this dissertation)
4. Minimum size of terminal node (*Nodesize*, 1 for this dissertation)

Random Forests produces the following output:

1. a classification tree generated using in-bag training data
2. the percentage of mis-classification using OOB data passing through the constructed tree (OOB classification error)
3. a ranking of each feature variable's importance measured using OOB data. The overall prediction error (OOB classification error) is the majority vote from all individual trained trees
4. vote and margin calculated for every LiDAR tree and class

The definition of the following terms will be discussed in further details as background information for the methods described in the following Chapters.

Importance: Importance is being assessed two different ways in RandomFoests, the first is called the mean decrease accuracy (MDA) and is calculated by estimating the OOB error that is recorded for each tree, e. Then, by random permutation, the f^{th} feature will produce a new OOB error e_f. Importance can be measured by calculating $e_f - e$, averaged over all trees and normalized by the standard deviation; this is called the mean decrease accuracy (MDA). If a feature has a large value for the mean decrease accuracy, it signifies that the feature is more important. This definition is chosen as a measurement of feature importance for this dissertation. The second means of measure for classification tree is obtained by the total decrease in node impurities (measured by Gini index) from splitting; averaged over all trees.

Vote: When each classification tree is generated, a subset of the feature is selected, let $X \subset \mathbb{R}^f$ be the features selected and y is the predicted class label. Where $y \in L$, L is the all the possible predicted labels. According to Schwing et al. (2011), the binary indicator variable for voting L can be written as $p_i(y|X)$. For a particular instance L, average vote can be calculated by summing all the votes for the particular class divided by the number of classifiers (T).

$$V(X) = \frac{1}{T}\sum_{i=1}^{T} p_i(y|X)\,, \text{where } V \subset \mathbb{R}^L \tag{2.9}$$

Margin: Margin is the distance from the data point to the decision boundary. In randomForest package of R, it is recorded as the proportion of the votes for the correct class minus the maximum proportion of the incorrect class. Recall y is the predicted class label. Let Y be the correct class label obtained from field validation. Margin can be written as the following:

$$MG(X) = \frac{1}{T}\sum_{i=1}^{T} p_i(y = Y|X) - \max_{y \neq Y}\left[\frac{1}{T}\sum_{i=1}^{T} p_i(y \neq Y|X)\right] \tag{2.10}$$

In this dissertation, Random Forests is applied in three different experiments:

1. In Chapter 3, 24 geometric descriptors (features) are used for the classification of tree genera, Random Forests is employed as the classifier
2. In Chapter 4, two different sets of features (geometric and vertical profile) forms two classifiers, both using Random Forests as base classifier and are ensemble (combined) to improve classification accuracies obtained from Chapter 3
3. In Chapter 5, the multi-class classification scheme from Chapter 3 is broken down into a series of binary classification. Each of the binary classifiers runs Random Forests

2.2 Description of study area and data

LiDAR data was collected on 7 August 2009, about 75 km east of Sault Ste. Marie, Ontario, Canada (Figure 2-1) by a Riegl LMS-Q560 scanner at an altitude of 122 to 250 m above ground level. Multiple flight passes was flown to ensure a combined pulse density of 40 pulses per m^2. The altitude for the power-line corridor site is lower as a higher point density is required for power line risk management, whereas the forested sites were acquired at a higher altitude. The point density of LiDAR data collected is approximately 40 pulses per m^2 with up to five returns per pulse.

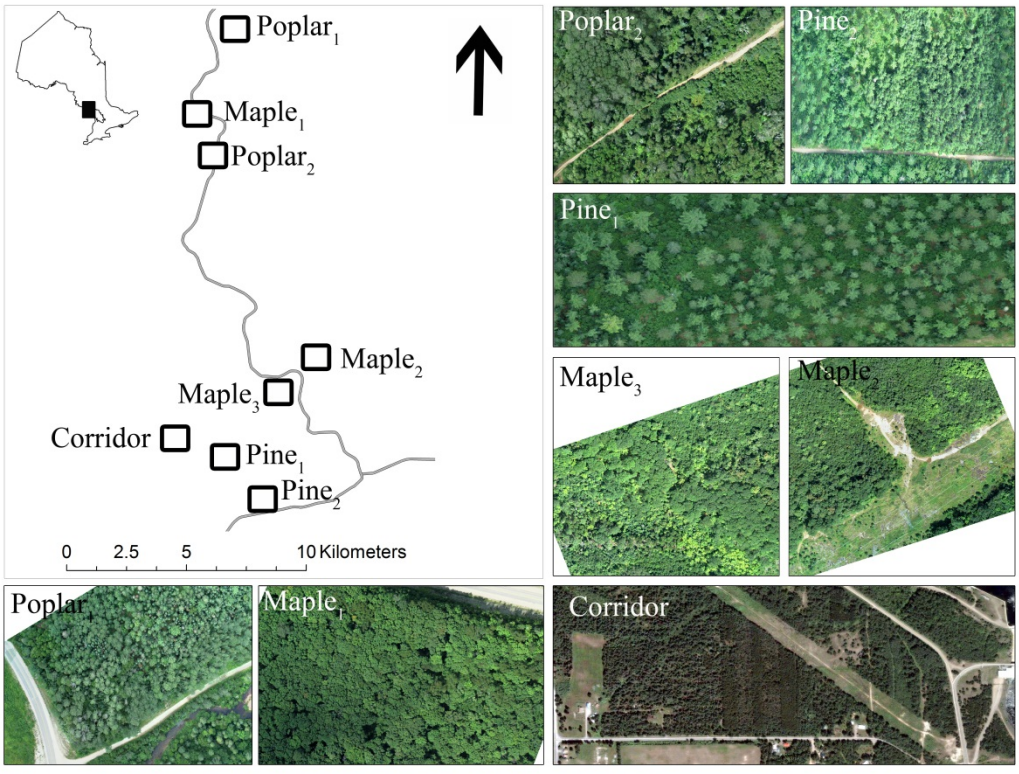

Figure 2-1 Map of the study area and the locations of the eight field survey sites.

Summer validation field surveys were conducted in 2009 and 2011. The eight field sites (Table 2-1) were selected to capture a diversity of environmental conditions and to sample the complexity of the region. One of the tasks for the first visit was to identify the tree species and measure tree attributes such as tree height, tree crown base height, tree crown diameter and diameter breast height (dbh) for validation purposes for the first field survey. The objective of the second survey is to record only tree location, species and dbh. The poplar sites (*Poplar*$_1$ and *Poplar*$_2$) are dominated by poplar trees and are situated in the northern part of the study area; *Poplar*$_1$ has substantially less understory vegetation than *Poplar*$_2$. The maple-dominated sites (*Maple*$_1$, *Maple*$_2$, and *Maple*$_3$) share very similar characteristics, as each site has a closed canopy and is interspersed with other deciduous species such as white birch and oak. The understory growth is vigorous with an abundant immature layer. The *Corridor* site is the most complex, which was selected especially to include trees that are difficult to identify.

Table 2-1 Plot distribution and characteristics of canopy structures.

Site	N. Trees	Species	Structure*	Cover	Tree Height (m)			Crown height (m)		
					Mean	Min	Max	Mean	Min	Max
Poplar$_1$	6	Po (6)	M, SS	> 65%	27.7	26.2	28.9	6.9	5.2	8.3
Poplar$_2$	40	Po (40)	D, MS	> 65%	25.0	20.0	27.1	7.5	5.4	9.8
Maple$_1$	20	Bw (8), Maple (12)	D, MS	> 65%	21.3	12.7	27.2	19.2	11.4	24.5
Maple$_2$	20	Bw (1), Maple (16), Oak (3)	D, MS	> 65%	21.1	16.5	27.8	19.0	14.9	25.0
Maple$_3$	6	Maple (6)	D, MS	> 65%	19.3	15.7	21.0	19.3	15.7	21.0
Corridor	48	Pw (12), Pr (4), Po (14), Bw (2), Pj (4), Sw (10), La (2)	D, MS	> 65%	19.9	2.2	28.5	13.3	2.2	24.7
Pine$_1$	40	Pw (17), Pr (23)	O	35%-65%	26.1	20.2	33.6	13.6	8.5	22.2
Pine$_2$	6	Pw (6)	M, SS	> 65%	24.9	22.3	27.1	12.8	10.3	16.8

* M = Moderate, D = Dense, O = Open, SS = Single Stratum, MS = Multiple Strata

Po = Poplar, Bw = white birch, Pw = white pine, Pr = red pine, Pj = jack pine, Sw = white spruce, La = larch

bracketed number in species column indicate the number of tree belong to the species

Individual tree crown isolation is difficult in this site due to partial occlusion by shadows, a high percentage of tree crown overlap with neighboring trees, and vigorous understory growth. In the *Corridor* site, growing conditions vary between the two sides of the transmission corridor due to differences in topography and sunlight penetration, which ultimately influence understory growth. The pine-dominated sites ($Pine_1$ and $Pine_2$) were selected to represent an open red and white pine canopy ($Pine_1$) and a white pine-dominated stand ($Pine_2$). Of the 186 trees sampled, 160 of them belong to one of the three genera *Pinus* (pine), *Populus* (poplar) or *Acer* (maple), which were selected to be considered in this paper not only because they represent the majority of the field validation, but also because they represent the dominant tree forms as depicted in Figure 2-2. Figure 2-2 shows an example of a LiDAR point cloud from each genera sampled at the northern Ontario field sites. Field validation identified species to include white birch (*Betula papyrifera* Marsh.), maple (*Acer saccharum* Marsh.), red oak (*Quercus rubra* L.), jack pine (*Pinus banksiana* Lamb.), poplar (*Populus* temuloides), white pine (*Pinus strobus* L.), white spruce (*Picea glauca* (Moench Voss)), and larch (*Larix laricina*).

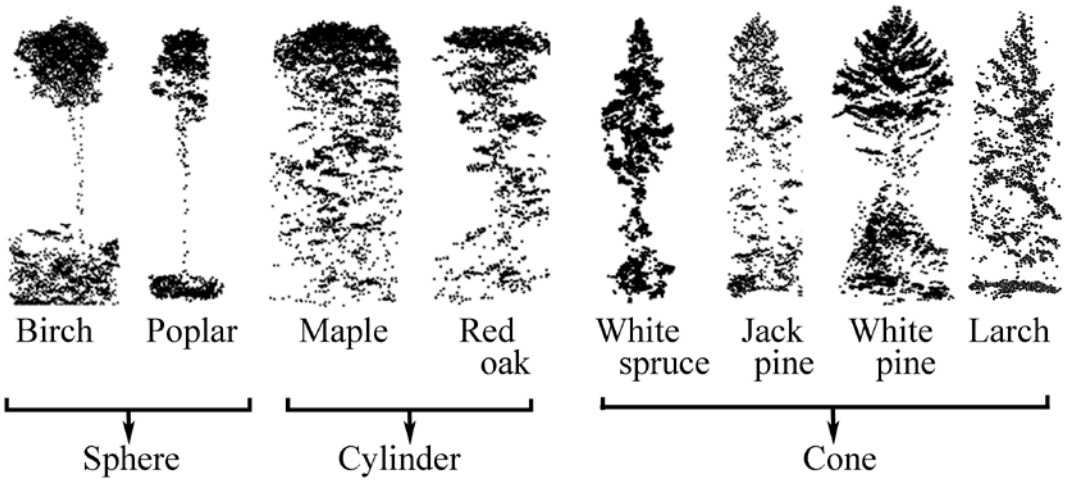

Figure 2-2 An example of LiDAR point cloud for each species identified in the field study and are grouped by three forms: 1) sphere, 2) cylinder, and 3) cone.

2.3 Preprocessing of the LiDAR data

There are two main preprocessing procedures; the first is the segmentation of LiDAR point cloud for individual trees and the removal of understory points for the isolation of tree crown points. The segmentation of individual tree is a manual process; the LiDAR trees are isolated prior to the field visits. The centres of the trees are estimated by the centroids of the isolated point clouds and are used for the verification of location in the field using a handheld GPS.

The understory is generally defined as the lowest layer of vegetation in an ecosystem, persisting below the main forest canopy, often representing a significant component of the total ecosystem biomass. However, the focus of this research is the geometry of tree crown and therefore the LiDAR points that belong to the understory is

removed before the extraction of geometric features. This is a semi-automatic process, an algorithm was written and was applied to each single tree. However, a visual check for each tree is performed and in the case where the algorithm does not match the visual check, visual interpretation was used to locate the tree crown base. The algorithm involve plotting a point density vertical profile for each tree, then, a best fit 4^{th} degree polynomial function was fit onto the profile. By solving for second derivative of the best fit function would locate the point of inflections. The higher point of inflection will be used as base of the tree crown and the lower point of inflection will be used as the top of the understory (not useful for this research).

All sampled trees (as LiDAR point clouds) undergo preprocessing that isolate individual tree crowns and remove understory points by analyzing point frequency distributions along height axes and the results from both processes are verified by visual interpretation. The understory points are removed prior to feature derivation because ecologically, understory points cause classification confusion and do not add to the characterization of tree crown interiors or outer surfaces. Trees that have distinct gaps between the overstory and the understory (bimodal distributions) can be segmented by a simple height threshold to eliminate points belonging to the lower group. Trees that have no distinct gap between the overstory and the understory had the understory removed manually by visual interpretation. This time-consuming procedure was typically required at the *Corridor*, $Maple_1$, and $Maple_2$ sites due to the abundance of immature saplings.

Chapter 3 Tree genera classification with geometric features

Published as: Ko, C., G. Sohn, and T.K. Remmel. 2013. Tree genera classification with geometric features from high-density airborne LiDAR. Canadian Journal of Remote Sensing 39(S1):S1-S13.[1]

Categorical recognition of a tree's genus is known to be valuable information for the effective management of forest inventories. This chapter will present a method for learning a discriminative model using Random Forests to classify individual trees into three genera: pine, poplar and maple. It is believed that both internal and external geometric characteristics of the tree crown are related to tree form and therefore useful in classifying trees to the genus level. The approach involves the extraction of both internal and external geometric features from a LiDAR point cloud because geometric features provide important information about the organization of the points inside the tree crown along with overall tree shape and form. This chapter originally developed 24 geometric features and then reduced the number of features in order to increase efficiency. These geometric characteristics, computed for 160 sampled trees from eight field sites, were

[1] I thank the journal, Canadian Journal of Remote Sensing, and my co-authors who have granted me the right to reproduce this article as a chapter in this dissertation.

classified using Random Forests and achieved an 88.3% average accuracy rate by using 25% (40 trees) of the data for training.

3.1 Introduction

The use of Light Detection and Ranging (LiDAR) has been useful for many applications and is gaining attraction for retrieving forest inventory attributes such as tree height and diameter (Hilker et al., 2008; Korpela et al., 2009). Among the many forest inventory attributes, tree genus (or species, if possible) is particularly valuable information to assist with estimating biomass, forest composition, wood volume, wildlife habitat, or disturbance effects (Bortolot, 2006; Niccolai, 2010; Suranto, 2009). In addition to forest inventory applications, genus information is often coupled with growth and yield tables or structural metrics of individual trees, which can facilitate the estimation of growth potential or leaning hazards of trees; therefore informing the scheduling of pruning or removal to ensure safe clearance zones along transmission line corridors (Jwa and Sohn, 2012).

The objective of this chapter is to investigate contributions of both internal and external geometric information about tree crowns, such as shape, branching, trunk structure, volumetric surface and height for genera classification, from discrete LiDAR point cloud data with high density (e.g. 40 pulses per m^2). This chapter categorized geometric features into two groups: external features which relate to the overall shape of the tree, and internal features which relate to the inner characteristics and branching structures. This chapter focuses on discussing methods deriving six important geometric

features from LiDAR point clouds through clustering and convex hull construction. These were selected from 24 geometric features by conducting feature reduction analysis in order to increase efficiency. Random Forests have been successfully studied for the prediction of forest inventory from LiDAR (Falkwski et al., 2009; Martinuzzi et al., 2009; Coulston et al., 2012; Yu et al., 2011). Random Forests predicts classification decision variables using ensembles of multiple classifiers, which often yield better classification results than single classifier (Na et al., 2010). Random Forests is a non-parametric discriminative classifier that allows us to effectively incorporate the internal features and external features for classification. This chapter explores potentials of Random Forests as a supervised learning classifier to predict three genera classes including maple, pine and poplar using LiDAR point clouds.

3.2 Methods

3.2.1 Overview of the methodology

To characterize genera that are believed to be ecologically different, this chapter rely on geometric features categorized into five broad families: 1) lines, 2) clusters, 3) volumes, 4) 3D buffers of points, and 5) overall tree height ratios. The geometric features characterize the spatial distribution of LiDAR points within the crowns of the individual trees which comprise the forest canopy. The line and cluster families of geometric features were selected to capture different elements of the primary linear structures within each tree crown. The geometric volume family characterizes tree crown volume, surface area of convex hulls, and alpha shapes, focusing on the outer surfaces of tree crowns,

while the 3D buffering family differentiates between clustered tree crowns and distributed LiDAR points. The final category (overall tree height ratios) considers the cloud of LiDAR points within individual crowns as a single object rather than segmentations relative to height profiles.

This chapter developed and computed 24 geometric feature summaries for each tree. A complete list of these features can be found in Table 3-1. The original 24 features were reduced to six according to their importance in order to increase the efficiency of classification by the Random Forests classifier. These six features are detailed in the following section. All sampled trees (as LiDAR point clouds) underwent preprocessing that isolated individual tree crowns and removed understory points by analyzing point frequency distributions along height axes.

The understory points were removed prior to feature derivation because they cause classification confusion and do not contribute to the characterization of tree crowns. Trees that have distinct gaps between the overstory and the understory were segmented by a simple height threshold to eliminate points belonging to the lower group. The understory was removed manually from trees that have no distinct gap. This time-consuming procedure was typically required at the Corridor, $Maple_1$, and $Maple_2$ sites due to the abundance of immature saplings; the automation of this procedure is complex and part of future research.

Table 3-1 List of derived 24 geometric features derived. (a) describes line related geometric features; (b) describes cluster related geometric features; (c) describes convex hull and alpha shape related geometric features; (d) describes 3D buffering related geometric features and (e) describes overall tree shape related geometric features. (Source: Ko. et al., 2012c)

Table 3-1 (a)

No.	Equation	Description
F1	$\frac{\sum_{n=1}^{k} L_n}{k \times H_t}$	Average line segment lengths divided by tree height
F2	$\frac{\sum_{n=1}^{k} L_n}{k \times H_c}$	Average line segment lengths divided by crown height
F3	$\frac{\sum_{n=1}^{k} L_n}{k} \times \frac{H_c}{H_t}$	Average line segment lengths multiplied by the ratio between tree and crown heights
F4	$\frac{\sum_{n=1}^{k} a_{xyn}}{k}$	Average line segment angles (rad) measured from the x-y plane to the line
F5	$\frac{\sum_{n=1}^{k} b_{xyn}}{k}$	Average line segment angles (rad) measured from the y axis to the line projected onto the x-y plane

where k = number of clusters at the end of the clustering algorithm, N_c = number of points in the crown, L_n = length of the line in the cluster, H_t = Tree height, H_c = Tree crown height, a_{xy}= angle between each line segment to xy plane, b_{xy} = angle between the projected line to y-axis

Table 3-1 (b)

No.	Equation	Description
F6	$(\sum_{n=1}^{k} \frac{N_n}{N_c}) \div k$	Average number of points in each cluster divided by the number of points in the crown
F7	$(\sum_{n=1}^{k} \frac{\sum_{l=1}^{N_n} d_l}{N_n}) \div k$	Average of the average orthogonal distance from each point to the line in the crown
F8	$(\sum_{n=1}^{k} \frac{\sum_{p=1}^{N_n} d_p}{N_n}) \div k$	Average of the average orthogonal distance from each point to the plane in the crown
F9	$(\text{F7} \div H_c) \times (\text{F8} \div H_c)$	F7 divided by the crown height multiplied by F8 divided by the crown height
F10	$(\sum_{n=1}^{k} \frac{V_n}{N_n}) \div k$	Average volume of the convex hull for each cluster divided by the number of points in the cluster

Where N_c = number of points in the crown, N_n = number of points in the cluster, d_l = orthogonal distance from each point to the line, d_p = orthogonal distance from each point to the plane, V_n = convex hull volume for cluster n, N_c = number of points in the crown.

Table 3-1 (c)

No.	Equation	Description
F11	$\frac{V_h - V_a}{V_h}$	Difference between the convex hull and alpha shape volumes compared to the convex hull volume
F12	$\frac{A_h - A_a}{A_h}$	Difference between the convex hull and alpha shape areas compared to the convex hull area
F13	$\frac{V_h}{N_c}$	Volume of the crown convex hull divided by the number of points within the crown
F14	$\frac{V_a}{N_c}$	Volume of the tree crown alpha shape divided by the number of points within the crown
F15	$\frac{\sum_{n=1}^{N_c} d_h}{N_c}$	Average distance from each point to the closest facet of the convex hull
F16	$\sqrt{\frac{\sum_{n=1}^{N_c}(d_h - \mathrm{F15})^2}{N_c}}$	Standard deviation of orthogonal distances from each point to the convex hull
F17	$\mathrm{F15} \div \mathrm{F16}$	Coefficient of variation

Where V_a = volume of the alpha shape of the tree crown, V_h= volume of the convex hull of the tree crown, A_a = area of the alpha shape of the tree crown, A_h = area of the convex hull of the tree crown, d_h = orthogonal distance form each point to the closest convex hull facet, N_c = number of points in the crown.

Table 3-1 (d)

No.	Equation	Description
F18	$\sum_{i=N_1}^{N_c} \sum_{j=N_1}^{N_c} V_{ij}$	Where i≠j; sum of overlapped volume between i^{th} and j^{th} spheres
F19	N_{ij}	Overlapped count of points captured by i^{th} and j^{th} spheres
F20	$\mathrm{F18} \div N_c$	Overlapped volume divided by the number of points within the crown
F21	$\mathrm{F19} \div {N_c}^2$	Count divided by the square of the number of points in the crown

Where Vij = overlapped volume between i^{th} and j^{th} speheres, N_{ij} = count of points captured by i^{th} and j^{th} spheres, N_c = number of points in the crown

Table 3-1 (e)

No.	Equation	Description
F22	$H_t \div \sqrt{\frac{A_{xy}}{\pi}}$	Tree height divided by the radius of the crown is circular when projected to x-y plane
F23	$H_c \div \sqrt{\frac{A_{xy}}{\pi}}$	Crown height divided by the radius of the crown is circular when projected to x-y plane
F24	$\frac{H_c}{H_t}$	Crown height divided by tree height

Where A_{xy} = area of tree crown projected to x-y plane

3.2.1.1 Feature reduction analysis

Feature reduction aims to produce a compact classification model by removing redundant features, whereby classification accuracy may decrease slightly, but at the benefit of a much simpler model and data requirement. Therefore, feature reduction is a trade-off between efficiency and accuracy, and it is often a subjective choice on how many features should be removed. To select six features from the original 24, first, correlations among features are examined so that highly correlated features can be removed to avoid issues of multi-collinearity. Five features are removed if feature correlation value is higher than pre-specified threshold of 0.85, which was empirically determined.

This simple threshold scheme facilitated eliminating obvious cases in feature correlation at the beginning of selection process; then further feature selection process concentrated on remained features. The features are sorted with mean decrease accuracy provided by Random Forests in descending order and analyzed the cumulative mean decrease accuracy for the 19 features. The rate of change in cumulative mean decrease accuracy measures the relative importance of a feature in relation to the rest of the features. The slope of the linear regression is steeper for the first 6 features (2.6) compared to the rest of the features (0.9) because relative importance for the first six features is higher from the rest of the features. To further support the decision, a method similar to Hudak et al. (2008) is applied, analogues to backward stepwise analysis: the classification was performed 20 times with the remaining set of 19 features and the

average classification accuracy was recorded. Features were removed one at a time followed by re-classification until a single feature remained. In general, it can be observed that the classification accuracy decreased when more features are being removed. The rate of change of classification accuracy is less for the first six features compared to the rest of the features. Smaller changes were observed during the removal of six important features because by removing one of these important features from the classification, there are other important features remained in the bin that is sustaining the high classification accuracy, therefore, further reduction is required. As the classification rate dropped more rapidly (steeper slope), that means this process have removed enough important features so that accuracy is dropping at a higher rate. It can be observed that from the dataset, the six most important features will lead to 8% classification reduction whereas the rest of the 13 (not as important) features will lead to a 43% accuracy reduction. From the above analyses, six features are retained for further classification purposes. To show the correlation between the selected six features, the pairwise correlation among the selected six features is listed in Table 3-2. These features are renamed to Features I, II, III, IV, V and VI for this paper.

Table 3-2 Pairwise correlation among the selected six features

	FI	FII	FIII	FIV	FV	FVI
FI	1.00	0.07	-0.01	0.04	0.02	0.03
FII	0.07	1.00	0.46	0.73	0.60	0.75
FIII	-0.01	0.46	1.00	0.23	0.58	0.40
FIV	0.04	0.73	0.23	1.00	0.42	0.58
FV	0.02	0.60	0.58	0.42	1.00	0.70
FVI	0.03	0.75	0.40	0.58	0.70	1.00

3.2.2 Selected feature derivation

The six selected features are described in Table 3-3. Features I, II and III relate to LiDAR point distributions within the crown (internal) and features IV, V, and VI relate to LiDAR point distributions at the outer surface of the crown (external). This dissertation defines external features as those related to the 3D boundary of interest and describes the shape of the crown. In contrast, internal features are those that describe the geometric structures within the defined boundary (clusters, lines, and planes that summarize the LiDAR point cloud). However, hard boundary for the crown is not defined, thus surface points and internal points are not mutually exclusive, meaning a particular point can be considered as an external point or internal point.

Table 3-3 List of the selected six features

No.	Equation	Description
I	$\frac{\sum_{n=1}^{k} L_n}{k \times H_t}$	Average line segment lengths divided by tree height
II	$\frac{\sum_{n=1}^{k} L_n}{k} \times \frac{H_c}{H_t}$	Average line segment lengths multiplied by the ratio between tree and crown heights
III	$\left(\sum_{i=N_1}^{N_c} \sum_{j=N_1}^{N_c} V_{ij} \right) \div N_c$	Overlap volume divided by the number of points within the crown
IV	$\frac{V_h}{N_c}$	Crown's convex hull volume divided by the number of points within the crown
V	$\frac{\sum_{n=1}^{N_c} d_h}{N_c}$	Average distance from each point to the closest facet of the convex hull
VI	$\frac{H_c}{H_t}$	Crown height divided by tree height

Where k = number of clusters at the end of the clustering algorithm, L_n = primary axis length of cluster, H_t = Tree height, H_c = Tree crown height, V_h= volume of the convex hull of the tree crown, N_c = number of points in the crown, d_h = orthogonal distance form each point to the closest convex hull facet, V_{ij} = overlapping volume between i^{th} and j^{th} spheres

3.2.2.1 Internal features (I, II and III)

The aim for Features I and II was to characterize the linear structures within tree crowns, which would require the construction and extraction of linear features evident within the LiDAR point clouds defining individual trees. Such extractions further necessitate the linking of nearby points, based on their association in space and in belonging to common elements within the crown. In order to derive the line and cluster related geometric features, a cluster merging-and-splitting algorithm is written in MATLAB®. The merging process involves over-segmenting given LiDAR points and then iteratively merging the segments to produce a coarsely clustered crown. The subsequent splitting process involves identifying the errors produced by the merging process due to over-merging, and accordingly splitting those clusters into an appropriate number by testing hypotheses until a final and optimally clustered crown is generated.

In the first part of the algorithm, the cluster merging process takes input from a LiDAR point cloud for a single tree, which is initially segmented into clusters using *k*-means clustering. The initial clustering partitions the individual tree crown into $k = 100$ clusters of LiDAR points. A 100 clusters tree (empirically selected) is chosen as a starting point because at this level, tree crowns were observed to be over-segmented; thus, points clearly belonging to a single branch were separated into multiple clusters. Then, pairs of these over-segmented clusters were merged if the candidate pair met certain conditions.

In order to determine whether two clusters should be merged, first, Euclidian distances between the *k* centroids of the over-segmented clusters are calculated. The centroids were then paired with the next closest centroid according to a computed distance. The centroid pairs were ranked based on this distance, with the closest distances being ranked highest. Additionally, the angle difference between the normal of the best-fit planes for each of the paired clusters is calculated. To estimate this angle difference, this algorithm adopted Principal Component Analysis (PCA) to approximate 3D planes for each cluster. PCA computes the principal components of each of the clusters within the tree; the first principal component defines the dominant direction of the cluster and is used as the direction vector in the following step. The second component is perpendicular to the first and is used as the normal for the plane. Planes for each cluster were defined by the normal vector and the centroid of the cluster, where *a* is the angle between the normal vectors of the paired planes. Cluster pairs with a lower *a* show a higher degree of surface normal similarity. Cluster pairs based on these two criteria (distance proximity and surface normal similarity) were ranked in order and used for a final hypothesis-validation process that determined the two clusters to be merged. In this step, each candidate pair was forced to merge, by which the original (prior to merging) and final (post merging) residuals (sum of distances from each point to the best fit plane) were calculated and the change (difference) was recorded. The change in residual is presented as a percentage from Equation (3.1), where R_o is the summation of residual distances prior to merging and R_n is the new summation of residual distances after merging. A small change

indicated that the pairs are likely to be coplanar and there was a higher chance that these two pairs belong to one bigger cluster.

$$\delta R = \frac{R_n - R_o}{R_o} \times 100\% \tag{3.1}$$

An optimal merging pair was determined by selecting the pair that has the smallest percentage change in Equation (3.1) among all potential pairs considered for merging within a predefined threshold. In this chapter the threshold was determined by using the 75^{th} percentile of the percentage change Equation (3.1) at k =100, such that the value varied among individual trees. The merging process continued until no candidate pair satisfied the aforementioned three conditions (proximity, surface normal similarity, potential merge test). The result from the clustered crown was used as an input for refining the splitting process.

As the second part of the algorithm, the splitting process served two purposes; the first was to identify clusters that were incorrectly merged (overly merged) and the second was to correct them. All clusters derived from the merging process produced the following five values for evaluation.

Let V_k be the convex hull volume for cluster k and N_k be the number of points in cluster k

$$\frac{V_k}{N_k} \tag{3.2}$$

As this value increase, it indicates that a cluster have a large volume relative to the number of points within the cluster; thus the cluster is likely better represented by more than one cluster.

Let d_p be the orthogonal distance from each point to the best fit plane and N_k be the number of points in cluster k.

$$\frac{\sum_{p=1}^{N_k} d_p}{N_k} \tag{3.3}$$

As this value increase, the cluster is likely non-planar.

Let d_l be the orthogonal distance from each point to the best-fit line and N_k be the number of points in cluster k.

$$\frac{\sum_{l=1}^{N_k} d_l}{N_k} \tag{3.4}$$

As this value increase, the cluster is likely non-linear.

Let $\vec{p} = (\hat{x}_k, \hat{y}_k, \hat{z}_k)$ *be the direction vector of the best-fit line passes through cluster k pointing away from the center of the point cloud*

$$\tan^{-1} \frac{\widehat{z_k}}{\sqrt{\hat{x}_k{}^2 + \hat{y}_k{}^2}} \in \left[0, \frac{\pi}{2}\right] \tag{3.5}$$

As this value increase, or closer to $\frac{\pi}{2}$, the line derived from the cluster is likely vertically oriented, which would be better represented by multiple lines, assuming that branching is typically more horizontal.

Let $\vec{q} = (\hat{x}_k, \hat{y}_k)$ *be the x and y component of* $\vec{p}$ *described in Equation (3.5);* $c = (\frac{\sum_{j=1}^{i} x_j}{i}, \frac{\sum_{j=1}^{i} y_j}{i})$ *and i represent the number of points in the given tree crown. Let* $\vec{r} = (c_x, c_y)$ *be the x and y component vector point from point c to mid-point of the segment* $\bar{q}$.

$$\cos^{-1} \frac{\vec{q} \cdot \vec{r}}{|\vec{q}||\vec{r}|} \in \left[0, \frac{\pi}{2}\right] \tag{3.6}$$

As this value increase, or closer to $\frac{\pi}{2}$, the derived line does not originate from the centre of the tree; instead, the line forms a skew line with the vertical, central axis of the crown.

Since Equation 3.2 to 3.6 have different ranges, to capture the potential commission errors from the merging step, each of the five criteria was used to select eight clusters with the highest values of each. The number of clusters was determined heuristically to achieve optimal performance. The total number of clusters needed to be considered in the splitting step equals to the union of all clusters was recorded. They will be assigned for testing in the next step.

The last step of the algorithm iteratively splits each of the identified clusters into an increasingly larger number of clusters (up to four clusters) by *k*-means clustering and the percentage change of residual distances from each point to the line is calculated using Equation (3.1). For any identified cluster, the residual distance will decrease as the number of splits increases. This is because fitting multiple lines through a fixed number of points will improve fitting, until an extreme case where each line consists of only two points; in this case, the residual distance will be 0. The changes in residual distances were recorded in each split and the optimal number of splits was determined by the maximum change in residual distances. A tree crown with multiple clusters was formed. Figure 3-1 is a flow chart that summarizes the merging-splitting algorithm for deriving the cluster, line and plane geometric features. Feature I records the average line segment lengths with

respect to the tree height and Feature II records the average line segment lengths with respect to the ratio between tree crown height and tree height.

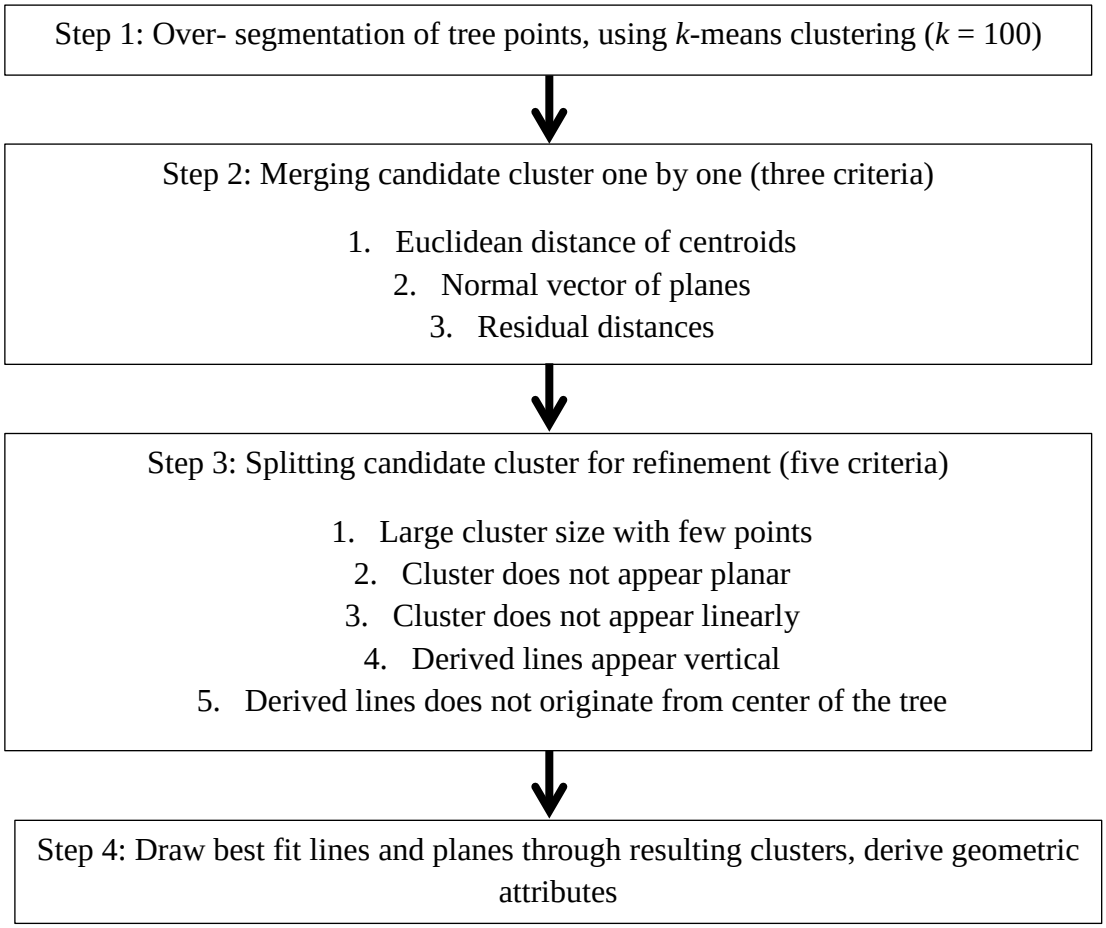

Figure 3-1 Flow chart summarizing the merging-splitting algorithm for deriving geometric features

Feature III quantifies the amount of clustering among points within the tree crown. In this section, the idea from Ko et al. (2009) was extended to three dimensions.

The points from the point cloud were buffered into spheres and the amount of overlapping volume between spheres was calculated. This information was used to study point proximity within crowns. Similar to 2D buffering, where circles intersect and overlap with a common area, three dimensional buffering has an intersection of spheres, where the intersection is measured as a common volume. If the spheres are not touching, the volume overlap will equal zero. If the point clouds are spatially clustered within the tree crown, the probability of each sphere overlapping each other increases, which will result a higher total overlap volume. The buffering distance was chosen to be 2% of the crown height to avoid scaling issues because if a fixed radius is chosen, small trees will end up having a relatively large sphere overestimating the volume of overlap. Feature III totals all of the overlap volumes and divides that value by the total number of points within the tree crown.

3.2.2.2 External features (IV, V and VI)

Features IV and V involve the calculation of a 3D convex hull, and the goal of these two features is to capture the geometric characteristics of tree crowns as a whole, rather than as segmented clusters. A convex hull is a convex 2D shape containing a set of points (Hornberg, 2006). In three dimensions, the convex hull of a tree crown becomes a solid convex volume formed by three-dimensional triangulations that contain all of the LiDAR points within the given tree crown. Feature IV calculates the total volume of the convex hull, dividing that value by the total number of enclosed points; this feature quantifies whether the crown has points that are tight and compact (low volume per the number of LiDAR points in each cluster) or loose and sparsely distributed clusters (high

volume per the number of LiDAR points in each cluster). Feature V provides the average distances from each point to the closest convex hull triangular facet. If all the LiDAR points lie on the facets (perfectly smooth crown), the value of this feature would be 0. However, if the LiDAR pulses penetrate deep inside the crown, the value of this distance will be large. The distance from each point to the closest triangular facet of the convex hull is

$$d_h = |\text{Proj}_{\vec{v}}\vec{w}| = \frac{|\vec{v}\cdot\vec{w}|}{|\vec{v}|} \tag{3.7}$$

where $\vec{v}$ is the normal vector of the closest triangular facet with a given LiDAR point, and $|\vec{v}| = 1$; $\vec{w}$ is the vector from the LiDAR point to one of the closest convex hull vertices. The distances were recorded in a column matrix and the sum of the distances was recorded. The distances from each point within the LiDAR cloud to the internal wall of the convex hull differentiates tree crowns based on their ability to permit penetration by LiDAR pulses. Feature VI outlines the shape of the tree as a whole object, describing the height of the crown relative to tree height.

3.2.3 Random Forests Classification

Random Forests is used to perform the classification of tree crowns. Random Forests is an ensemble classifier comprised of many classification trees for categorical data, or regression trees for continuous data trained from a subset of the data (Breiman,

2001, Liaw and Wineder, 2002). Unlike classification tree or regression tree analysis, where a single classification tree is produced for classification, Random Forests constructs a user-defined number of trees, *Ntree*, by using randomly selected feature variables. For each iteration of classification tree construction, a subset of the data is drawn by sampling with replacement (a bootstrap sample), where about ⅓ of the data (out-of-bag data or OOB data) is set aside for testing the classification accuracy of the current tree and ⅔ of the data (in-bag data) is used to construct a single decision tree. In Random Forests, the decision tree grows by consequently splitting a parent node into binary descendent nodes, by using the random subset of features. The best split at each node is defined by an optimal feature subset that minimizes the class impurity within the descendant nodes.

Random Forests classification has successfully been employed in forestry applications. For example, Falkowski et al., (2009) used 34 features derived from LiDAR, then reduced the number to 9 to classify successional forest stages. Martinuzzi et al., (2009) used Random Forests to classify snag types by deriving 34 LiDAR features (19 of them related to the canopy height point distribution and 15 of them related to the topographic information). The implementation of classification algorithm is completed within the *randomForest* package for R (R Development Core Team, 2011; Breiman, 2001). The main input parameters of Random Forests for classification include: 1) training sample labeled with known species and geometric features, 2) the number of feature variables randomly sampled at each split (*mtry*, 3 for this research paper), 3) the number of trees generated within each iteration (*Ntree*, 1000 for this research paper) and

4) Minimum size of terminal node (*Nodesize*, 1 for this research paper). Random Forests produces the following output: 1) a classification tree generated using in-bag training data, 2) the percentage of mis-classification using OOB data passing through the constructed tree (OOB classification error), and 3) a ranking of each feature variable's importance measured using OOB data. The overall prediction error (OOB classification error) is the majority vote from all individual trained trees.

The importance of each feature can be calculated by estimating the OOB error that is recorded for each tree, e. Then, for each feature, f_f, where f=1…6, the randomly permuted f^{th} feature is used and therefore will produce a new OOB error e_f. Importance can be measured by calculating $e_f - e$, averaged over all trees and normalized by the standard deviation; this is called the mean decrease accuracy. If a feature has a large value for the mean decrease accuracy, it signifies that the feature is more important in classifying the tree genera.

When considering the optimal amount of data for training for any supervised classification, one need to consider the error caused by changing the training size. The bias of classification measures the quality of match, and the variance measures the precision of match; both can be lowered with a large training size (Duda and Harts, 2000). Unfortunately, the reality of having a large training size is impractical because it would involve abundant ground validation data that is prohibitively expensive to collect. In order to select the optimal percentage for segregation training and validation data while running Random Forests, the classification accuracy is tested with an increasing proportion of training data. Starting with 5% of the data for training (95% for validating),

the training data increment in 5% steps until it reached 95%; at each partition, Random Forests is ran 20 times.

3.3 Results and discussion

3.3.1 Selected features

Features I and II are the average derived line lengths of the tree height and the ratio between crown height and tree height respectively. They are both obtained from the results from the merging-and-splitting algorithm. Figure 3-2 shows an example of a pine tree (it is found that pine trees have the most organized structure and therefore use them for illustation). Figure 3-2(a) shows a pine tree with clustering completed manually by visual analysis, where lines on the figure represent the best-fit lines through each cluster. Figure 3-2(b) shows the same pine tree with lines derived after the merging process (merging in adherance to the three defined conditions). It is evident that some lines are vertically oriented and are thus representative of clusters that are overly merged. Figure 3-2(c) shows the same pine tree after running both the merging and splitting algorithms. These two features are designed to separate trees that have long line lengths (larger cluster size) with short line lengths relative to the tree height and crown height. Maples have relatively longer line segments due to their large crowns, while poplars tend to have shorter line segments due to their smaller crowns.

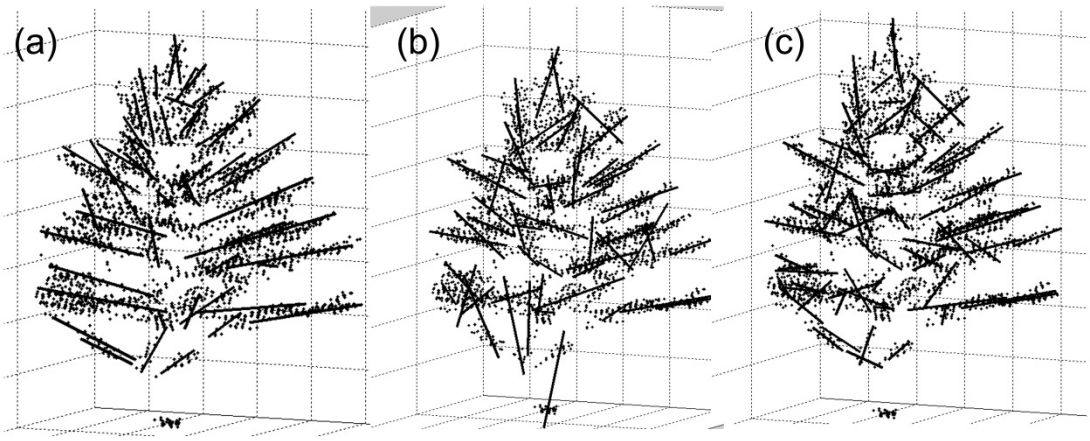

Figure 3-2(a) shows the result of a manually clustered conical tree; (b) shows the result of the lines derived from the merging process and (c) shows the result of the lines derived from the splitting process.

The 3D buffering feature (Feature III) is designed to distinguish trees that have clusters detected inside their crowns from those trees that do not. This feature highlights the spatial distribution of points by clumping points that are close together and it is important for the analysis of the spatial relationship (proximal spacing) among the points; the density of the points within a given volume plays a substantial role in determining the general tree form, and hence genera and perhaps species. Pine trees have layered branching structures, resulting in LiDAR points that are located as compact, horizontally layered clusters mimicking the branching, whereas poplars with spherical crowns have LiDAR points that are more evenly distributed within the volume of the crown, causing a lower value for this feature. Figure 3-3 shows the result for this feature that sums all the

volumes from overlapping LiDAR points if each point were buffered outward by a distance equaling 2% of the crown height.

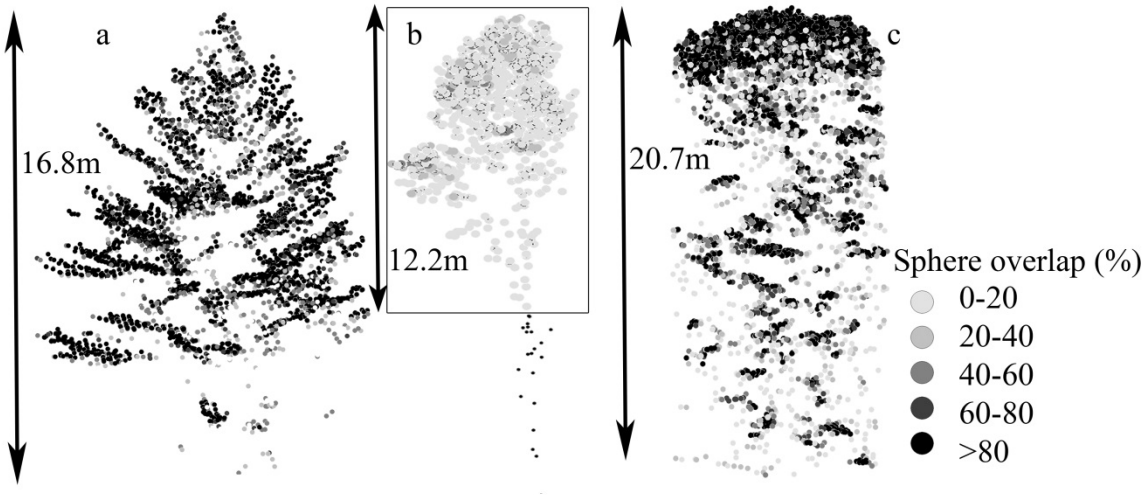

Figure 3-3 Example (a) pine tree, (b) poplar tree, and (c) maple tree; all crown points buffered outward by a distance equal to 2% of the crown height.

Feature IV is designed to quantify the point density of a given crown (i.e., how much space each point occupies within a crown); poplars tends to have smaller values whereas maples have the largest values. This is explained by the fact that most maple samples have a high density of points at the crown top and center portion of the tree, and less at the bottom and the sides, resulting in an underestimation of tree crown volume when constructing convex hull. Moreover, there is a lot of crown overlap with neighboring trees; therefore, single tree delineation may sometimes only consider the

central portion of the maple tree, resulting in an overestimation of crown point density. Conversely, poplars have smaller crowns but with LiDAR points more evenly distributed throughout. The disadvantage of this feature (compared to buffering) is that it averages the point density over the entire crown, so localized clusters (portion of the tree crown with high point density) are not well represented.

Feature V is designed to differentiate crowns that permit LiDAR signals to reach to the inside of the crown from the tree crowns that do not. Large values indicate points that are far away from the convex hull wall (more points are deep inside the crown). Pine has the highest value because pine tree crowns are more open and are comprised of smaller leaves, thus more points are recorded within the crown. Feature VI calculates the ratio between the crown and tree heights. Figure 3-4 shows a schematic diagram of the pine, poplar, and maple trees; box plots for tree and crown heights are included to show the distribution of tree size sample of the dataset. Maples have a ratio close to one and poplars tend to have the smallest ratio.

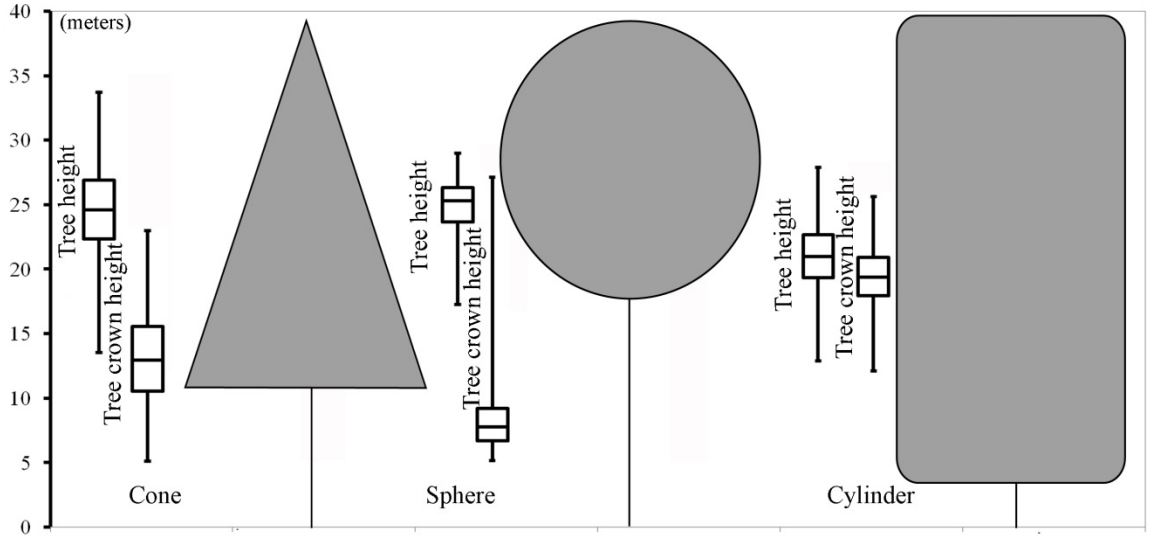

Figure 3-4 A schematic describing the components of Feature VI; box plots for tree height and crown height for each tree shape is included to show the distribution of tree size sample of the dataset.

3.3.2 Random Forests results

An optimal training set size is tested to maintain an effective classification by randomly partitioning the 186 tree samples into training and validation sets in varying proportions, using the six selected features. At each partition, Random Forests is performed 20 times and mean classification accuracies were recorded as indicators of classifier performance. Figure 3-5 shows the learning curve presenting the mean classification accuracy versus the proportion of training samples relative to those used for validation. The mean validation accuracy does not change much with a larger training

proportion above ~25%. The figure also shows that by using only 5% of the data for training, the accuracy already approaches 74%. In this chapter, 25% of the data is designated to run Random Forests and 75% of the data to validate the results.

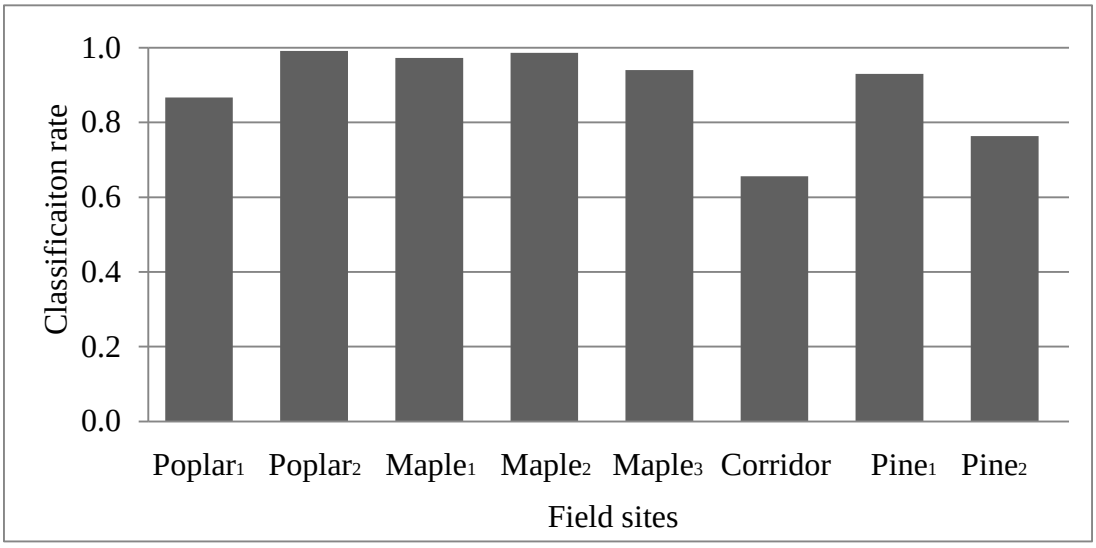

Figure 3-5 The average change in classification accuracy with increasing proportion of training data. Bars represent the minimum and maximum accuracy over 20 iterations.

When using a 25:75 percent ratio of training to validation data, the average classification accuracy was 88.3% over 20 iterations (Table 3-4) shows a confusion matrix for the validation accuracy averaged over 20 iterations. Figure 3-6 depicts the mean decrease in accuracy for the six selected features; Feature VI is the most important and indicates the importance of visualizing a tree as a whole object. Although the density of point distribution decreases with the depth of penetration within the tree crown, Yu et al. (2011) demonstrated that the most important features for predicting tree attributes are

situated near the top of the tree. It is nevertheless important to consider all points within the tree structure because the overall shape of the tree plays a significant role in classification. The second most important feature is Feature III, 3D buffering, and this shows the importance of inspecting information internal to the crown. Internal geometric information from the crown is difficult to obtain by methods of data acquisition other than high density LiDAR data, which allowed the derivation of such information. Of the six selected features from the entire set of 24, three features are related to internal crown geometry (I, II and III), further demonstrating the significance of internal geometry for tree genera classification.

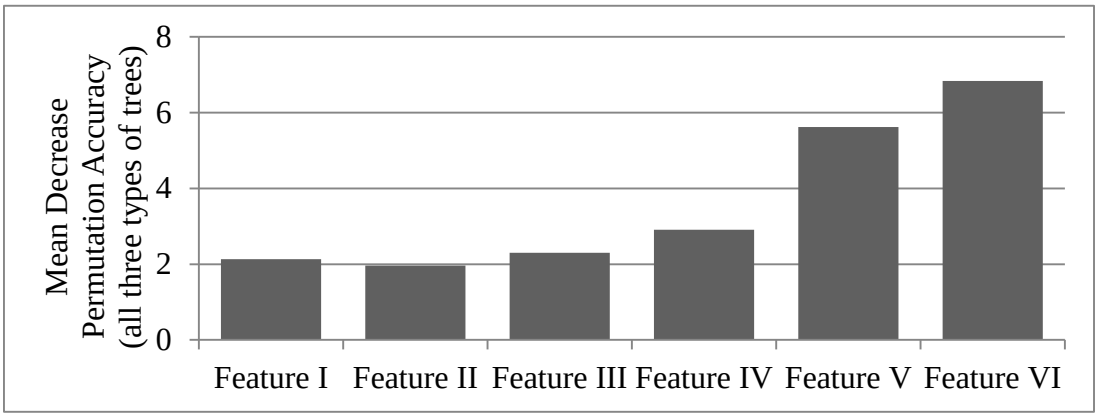

Figure 3-6 The mean decrease permutation accuracy for selected 6 features.

Table 3-4 Confusion matrix for the validation accuracy averaged over 20 iterations using 25% of the data for training using Random Forests.

		Expected			
		Pine	Poplar	Maple	User's Accuracy (%)
Predicted	Pine	43.0	6.7	0.7	85.5
	Poplar	5.0	37.8	0	88.4
	Maple	1.7	0.2	25.1	93.1
	Producer's Accuracy (%)	86.6	84.8	97.5	

3.3.3 Classification performance for individual site

Figure 3-7 shows the classification rate for individual sampling sites using 25% of the data for training and 75% of the data for validation. The classification rate was calculated as the number of times a tree is correctly classified during the 20 iterations. $Poplar_2$ had the highest classification accuracy (close to 100%) because spherical trees alike to poplars are the most unique in overall tree form, where the size of the tree crown is smallest and the spatial point distribution within the tree crown is even, relative to the crown height. The poplars at $Poplar_2$ are well separated, as they are open canopy sites with minimal understory growth; thus, tree detection and crown isolation are simplified and improve classification accuracy. $Pine_1$, $Pine_2$, $Maple_1$, $Maple_2$, and $Maple_3$ have classification rates ranging from 76% ($Pine_2$) to 98% ($Maple_2$). The *Corridor* site had the lowest classification rate (66%), as trees on that site represent more heterogeneous

environmental conditions than other sites (shadowing, high density of stems, intertwined crowns, tall and dense understory).

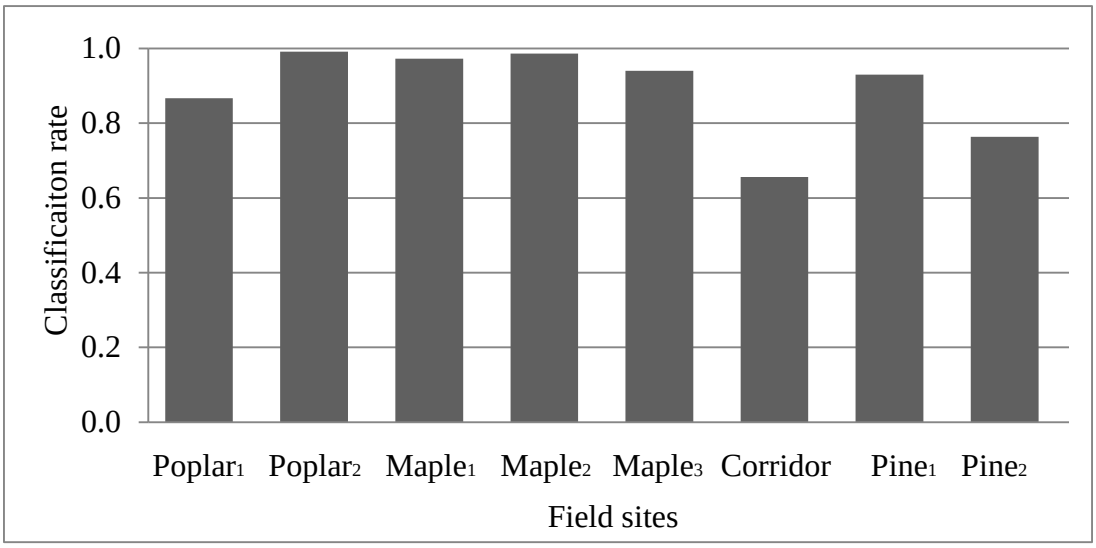

Figure 3-7 Classification rate for individual field sites averaged over 20 iterations.

3.4 Conclusions

In this chapter, the possibility of using geometric features to classify trees into three genera with a perspective different from the traditional vertical profile approach is examined. New descriptors for point cloud processing designed for tree classification applications are developed. Flying at low altitudes allowed LiDAR signals to better penetrate the tree canopy, providing a dataset with high point density and the possibility of deriving geometric features representative of crown interiors. Six geometric features are identified, which help to efficiently classify tree genera. Of all features, the overall

shape of the tree plays a significant role in identification, meaning that it is important to interpret trees as whole objects. As a result, a precise single-tree segmentation algorithm is important for classification accuracy. In this chapter, a manual approach is chosen to isolate single trees within the LiDAR point cloud to test the performance of the geometric features. However, for operational applications, automatic tree delineation will have to be implemented.

3D buffering of features was the second-most important feature, and was particularly effective in highlighting the tree branching structures observed in coniferous trees (Figure 3-3**.**). Random Forests is performed for the classification and it is found that by using only 5% of the data for training, 74% classification accuracy can be achieved. The average accuracy over 20 iterations when using 25% of the data for training was 88.3%. Additionally, Random Forests produced feature importance measurements that allowed us to gain knowledge about the effectiveness of each feature for genus identification. Three selected features are associated with the internal structures of the crown and three related to the overall tree crown. This project selected 160 trees from eight field sites, taking local environmental variations in considerations. The classification rate for *Corridor* was lower than for $Poplar_1$, $Poplar_2$, $Maple_1$, $Maple_2$, $Maple_3$, $Pine_1$ and $Pine_2$ because of the heterogeneous environmental conditions. It is believe that the classification accuracy is related to the quality of the segmentation of LiDAR point cloud, such as the completeness of the tree crown (no shadow), effective removal of understory, and minimally intertwined crowns. The classification accuracy for

any site that is near or on the transmission lines, with good quality single-tree segmentation, will yield a high classification rate.

Further investigation is necessary in order to successfully classify trees from the genus, to the species level. This would require having substantially more samples as well as additional features and classes to train the Random Forests algorithm. This chapter also emphasizes the importance of having high quality LiDAR data that includes a sufficient point density, non-occluded crowns, and crowns are spatially separated to permit easy extraction of individual trees. While this may allow the development of high quality exemplars of tree genera, in reality, the development of second-generation algorithms will need to deal with closed canopy situations. Finally, techniques for understory removal or suppression are warranted to yield accurate measurements of crown base heights for overall tree shape analysis.

Chapter 4 Ensemble classification of tree genera

This chapter will discuss an ensemble method for the classification of tree genus using LiDAR (Light Detection and Ranging) data. The two classifiers use different sets of features: 1) features derived from geometric information, and 2) features derived from vertical profiles using Random Forests as the base classifier. The result shows that classification accuracy can be improved by combining these two classifiers compared with classification using an individual classifier. An additional point-density analysis is performed to study the influence of decreased point density to classification accuracy results. The training genera include pine, poplar, and maple within a study area located north of Thessalon, Ontario, Canada. The average classification accuracy for the geometric classifier and vertical profile classifier are 88.0% and 88.8%, respectively, with improvement to 91.2% using the ensemble method.

4.1 Introduction

The ability of airborne LiDAR to acquire 3D information has spawned many forestry applications, for example: Holmgren and Persson (2004); Barilotti et al. (2009); Brandtberg (2007); Kato et al. (2009); Ørka et al. (2007, 2009); Vauhkonen (2008, 2009, 2010); Korpela et al. (2010) and Kim et al. (2011). The objective of this chapter is to discuss and apply classification using an ensemble method using two types of features. An ensemble method in classification is the training of multiple classifiers to solve the same problem. Ensemble classification in different disciplines has been called committee-based learning, multiple classifier systems, or the mixture of experts

(Samadzadegan et al., 2010; Ruta and Gabrys, 2000; Zhou 2012). The major criterion for an effective ensemble system is to provide an increase in classification accuracy (Ali and Pazzani, 1996; Breiman 1998 and Kavzoglu and Colkesen 2013); examples of such studies in remote sensing include Samadzadegan et al. (2010); Liaw and Wiener (2002); Kumar et al. (2002) and Kavzoglu and Colkesen (2013).

The first set of features used for classification derived from the geometry of the LiDAR point distribution reflected from the tree. Previous studies with a similar approach for obtaining tree species/genera metrics include Barilotti et al. (2009) and Kato et al. (2009), both fitting curved surfaces to the individual LiDAR tree crowns to obtain characteristic shape metrics. Vauhkonen (2008, 2009, 2010) compute the alpha shapes of tree crowns and obtain metrics from the shapes for tree species classification; in particular, Vauhkonen (2010) showed a classification rate of 78% classifying Scot pine, Norway spruce and deciduous trees.

The second set of features used derived from the vertical profile of the reflected LiDAR points, including the statistics summarizing the point distribution within specific height percentiles or the entire profile. Examples of research using the vertical distribution of height and/or intensity include Holmgren and Persson (2004) who classified Norway spruce and Scots pine, achieving an accuracy of 95%, Brandtberg (2007) classified oak, red maple and yellow poplar with an accuracy of 64%; Ørka et al. (2007, 2009) achieved an accuracy of 74% for classifying spruce, birch and aspen, and 88% for classifying large Norway spruce and birch trees respectively. Korpela (2010) achieved an accuracy of up to 90% classifying Scots pine; and Kim et al. (2011)

distinguish 8 deciduous from 7 coniferous genera with up to 74.9% classification accuracy.

In the previous chapter (Chapter 3), 24 features were derived based on geometric information; a full list of geometric and vertical profile features are shown in Table 3-1 and Table 4-1 Subsequently the number of features were reduced to 6 and 26 respectively and an ensemble method was introduced that combines the two classifiers and improves classification results. Vauhkonen et al. (2008) Magnusson et al. (2007) also demonstrated that the accuracy in estimating tree attributes decreases when pulse density decreases; thus an additional density sensitivity analysis was performed to assess the lower limit for the suggested classification scheme.

4.2 Methods

Geometric features were derived by clustering LiDAR point clouds that represent individual trees. The merging-splitting algorithm that groups representative points (belonging to a single branch) into a common cluster is described in Chapter 3. Best-fit lines passing through each cluster centroid are drawn and the characteristics of those lines are calculated. The feautres also include metrics related to the convex hull of the LiDAR point cloud and 3D buffering of individual points.

Table 4-1 Summary of features derived from vertical point profile: F= first; S= single; L= last; SD= standard deviation; CV = coefficient of variation, source: Ko et al. (2012a)

10^{th} percentile	50^{th} Percentile	90^{th} percentile
% of canopy return ($V1_F$, $V2_S$, $V3_L$)		
% return count ($V4_F$, $V5_S$, $V6_L$)	% return count ($V7_F$, $V8_S$, $V9_L$)	% return count($V10_F$, $V11_S$, $V12_L$)
Mean height ($V13_F$, $V14_S$, $V15_L$)	Mean height ($V16_F$, $V17_S$, $V18_L$)	Mean height ($V19_F$, $V20_S$, $V21_L$)
Mean height of canopy return ($V22_F$, $V23_S$, $V24_L$)		
SD of height ($V25_F$, $V26_S$, $V27_L$)	SD of height ($V28_F$, $V29_S$, $V30_L$)	SD of height ($V31_F$, $V32_S$, $V33_L$)
SD height For canopy return ($V34_F$, $V35_S$, $V36_L$)		
CV height For canopy return ($V37_F$, $V38_S$, $V39_L$)		
Kurtosis of variation height For canopy return ($V40_F$, $V41_S$, $V42_L$)		
Skewness of variation height For canopy return ($V43_F$, $V44_S$, $V45_L$)		
Mean intensity ($V46_F$, $V47_S$, $V48_L$)	Mean intensity ($V49_F$, $V50_S$, $V51_L$)	Mean intensity ($V52_F$, $V53_S$, $V54_L$)
Mean intensity of canopy return ($V55_F$, $V56_S$, $V57_L$)		
SD of intensity ($V58_F$, $V59_S$, $V60_L$)	SD of intensity ($V61_F$, $V62_S$, $V63_L$)	SD of intensity ($V64_F$, $V65_S$, $V66_L$)
SD intensity of canopy return ($V67_F$, $V68_S$, $V69_L$)		
CV intensity of canopy return ($V70_F$, $V71_S$, $V72_L$)		
Kurtosis of variation intensity of canopy return ($V73_F$, $V74_S$, $V75_L$)		
Skewness of variation intensity of canopy return ($V76_F$, $V77_S$, $V78_L$)		

The second set of features (height attributes and intensity attributes) summarizes the properties of vertical point distribution within each tree crown, including the mean, standard deviation, coefficient of variation, kurtosis and skewness of the height and intensity distribution for the entire crown. There are 78 features for each tree derived. Table 4-1 summarizes these features. Each tree is height normalized and segmented into 10 vertical slices. The 10^{th} percentile features represent the LiDAR points belonging to the bottom 10^{th} percentile of the tree crown height whereas the 90^{th} percentile features represents the points located at the top of the tree. Features include “first of many” returns; “single return” and “last of many” returns. Feature numbers are in bold. V1 to V12 are attributes related to counting the percentage of points belonging to the first, single, or last category in different percentiles. V13 to V45 are features related to statistics of height and V46 to V78 are features related to statistics of un-calibrated intensity values. Figure 4-1 shows an example tree with its vertical point distribution profile from each genus of interest, Figure 4-1 (a), Figure 4-1 (b), Figure 4-1 (c) are an examples of pine, poplar and maple respectively.

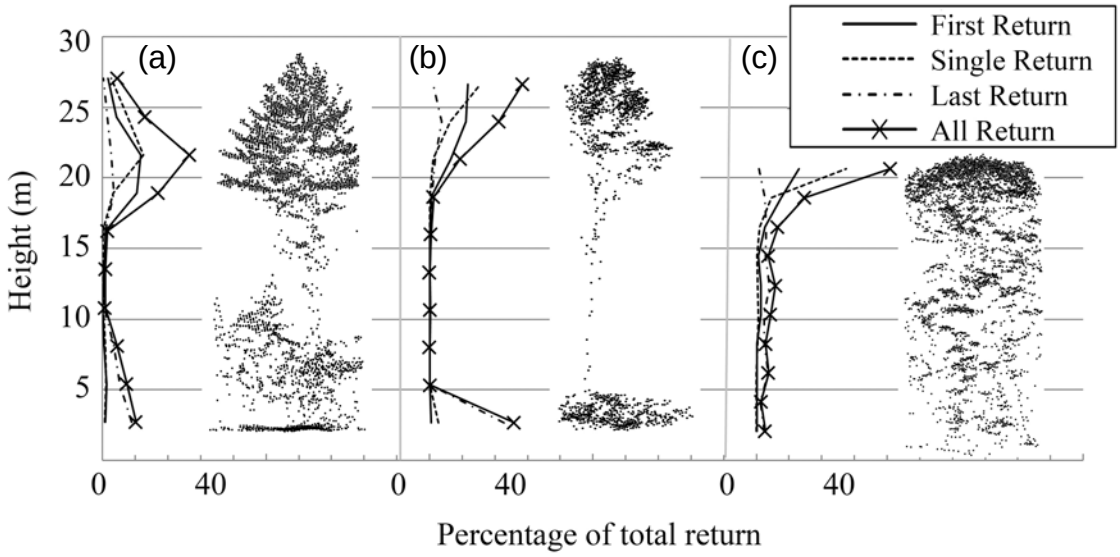

Figure 4-1 LiDAR data for pine (a), poplar (b) and maple (c) examples together with their corresponding vertical point distributions. source: Ko et al. (2012b)

4.2.1 Random Forests

Random Forests itself is an ensemble classifier; it combines many classifications (categorical data) or regression trees (continuous data) for making a final labeling decision (class labels) (Breiman 2001; Liaw and Wiener, 2002). The classification algorithm was implemented within the randomForest package for R (R Development Core Team, 2013; Breiman 2001). In Random Forests, about 37% of training data is partitioned for estimating the classification accuracy and it is called the out-of-bag (OOB) data, whereas using the rest of the data (in bag) for tree construction. The importance of each feature can be calculated by first, estimating the OOB error from a classification

tree, *e*. Second, by random permutation, the f^{th} feature will produce a new OOB error e_f. Third, the mean decrease accuracy (MDA) value for each feature can be calculated by e_f – *e*. This value is averaged over all trees and normalized by the standard deviation. If a feature has large mean decrease accuracy, it is a more important feature. Additionally, the proportion of vote and margin for each LiDAR tree are calculated and recorded.

Let $X \subset \mathbb{R}^f$ be the features selected for classification, y be a predicted class label such that $y \in L$. According to Schwing et al. (2011), binary indicator variable for voting the L instances can be written as $p_i(y|X)$, therefore the average vote for a LiDAR tree to be assigned as one of the classes with given *X* can be defined as (4.1), the summation of all votes for the particular class divided by the number of classifiers (*T*) that make this decision:

$$V(X) = \frac{1}{T}\sum_{i=1}^{T} p_i(y|X) \text{ , where } V \subset \mathbb{R}^L \tag{4.1}$$

The final label (y*) is decided by the majority voting scheme over *T* base classifiers described by (3.1):

$$y^* = MV(X) = \text{argmax}_{y \in L} \frac{1}{T}\sum_{i=1}^{T} p_i(y|X) \tag{4.2}$$

For the partitioned training data, one can obtain margin, described by Brieman 2001 as:

$$MG(X) = \frac{1}{T}\sum_{i=1}^{T} p_i(y = Y|X) - \max_{y \neq Y}\left[\frac{1}{T}\sum_{i=1}^{T} p_i(y \neq Y|X)\right] \quad (4.3)$$

For the partitioned validation data, although the field-validated Y is known, however, this section will treat it as unknown for estimating the prediction accuracy. Assume the correct class label is the maximum proportion of the vote; therefore, the first term of the margin is replaced by (4.2) and the pseudo-margin (*PG*) for the validation from the classifiers becomes:

$$PG(X) = \max_{y \in L}\frac{1}{T}\sum_{i=1}^{T} p_i(y|X) - \text{second\_max}_{y \in L}\frac{1}{T}\sum_{i=1}^{T} p_i(y|X) \quad (4.4)$$

The margin is defined as the distance from the data point to the decision boundary. In the randomForest package of R, margin is recorded as the proportion of votes for the correct class minus the maximum proportion of the incorrect classes. The larger the value $MG(X)$, the more confident is the correctness of the classification. For prediction purposes, margin is calculated by assuming the correct prediction equals to the class label from the maximum vote, denoted by $PG(X)$. The larger the value, the more confident can be placed in the prediction and this value will be used to filter out LiDAR trees that potentially exhibit incorrect classification from the base classifier. The main input parameters of Random Forests include: 1) labeled training samples, 2) the number of feature variables randomly sampled at each split (*mtry* = 2 for the geometric classifier and 5 for the vertical profile classifier), 3) the number of trees generated within each iteration (*Ntree* = 1000 for this example) and 4) minimum size of terminal nodes

(*Nodesize* = 1 for this example). Random Forests produces the following output: 1) a ranking of each feature variable's importance calculated by mean decrease accuracy measured using (OOB) data, 2) a randomForest classification object for testing the validation data, and 3) vote and margin calculated for every LiDAR tree and class.

4.2.2 Feature reduction

Feature reduction for each classifier is performed in two steps. The first is to remove the highly correlated features to avoid issues of multi-collinearity. This is done by calculating the pairwise correlation table and then removing features with $r > 0.85$, an empirically determined threshold. The second step is performed by Sequential Backward Selection (SBS) (Jain and Zongker, 1997 and Serpico et al., 2002). In SBS, a user-defined objective function J is needed to assess the performance of the feature subset; the optimal subset of features can be chosen by removing one least important feature at a time, starting with the full feature set until a single feature remains. By either maximizing or minimizing the objective function, the optimal number of features can be determined. This framework is applied with MDA values obtained from randomForests.

First, the cumulative MDA values are calculated and ranked in descending order for both classifiers, then plotted against the number of features being removed. Both graphs (one for geometric features and one for vertical profile features) show that cumulative MDA values decrease as the number of features removed increases. The rate of decrease increases because the important features are being removed later. The objective function is derived from fitting two linear functions through each curve. The first line will regress through cumulative MDA values that are being removed from SBS

and the second line will regress through the cumulative MDA values remaining. The rate of change (slope) measures the relative importance of the removed feature with respect to the rest of the features; the optimum fitting is then found by minimizing the residual sum of squares from each point to the lines Equation (4.5). The framework of the feature reduction analysis for geometric and vertical profile classifier is shown in Figure 4-2 as a flow chart. The result of the analysis for the geometric classifier is shown in Figure 4-3 (a) and result of the vertical profile classifier is shown in Figure 4-3 (b).

Let $MDA_{1...}MDA_f$ be the mean decrease accuracy calculated from Random Forests for features, $f_1 \ldots f_f$ ranked in descending order, the cumulative *MDA* plot for geometric classifier can be found in Figure 2 and denoted as *A(n)*, where *n* is the number of features removed. *f* is the total number of features available (19 in this example, reduced from 24 by the removal of highly correlated features).

Let l_i be the linear function through the 1^{st} to n^{th} values and l_j be the linear function through the n^{th} to f^{th} values such that P_i and P_j be the predicted *MDA* for the best fit line l_i and l_j, respectively.

$$J = min\left(\sum_{i=1}^{n}(P_i - A(i))^2 + \sum_{j=n}^{f}\left(P_j - A(j)\right)^2\right) \tag{4.5}$$

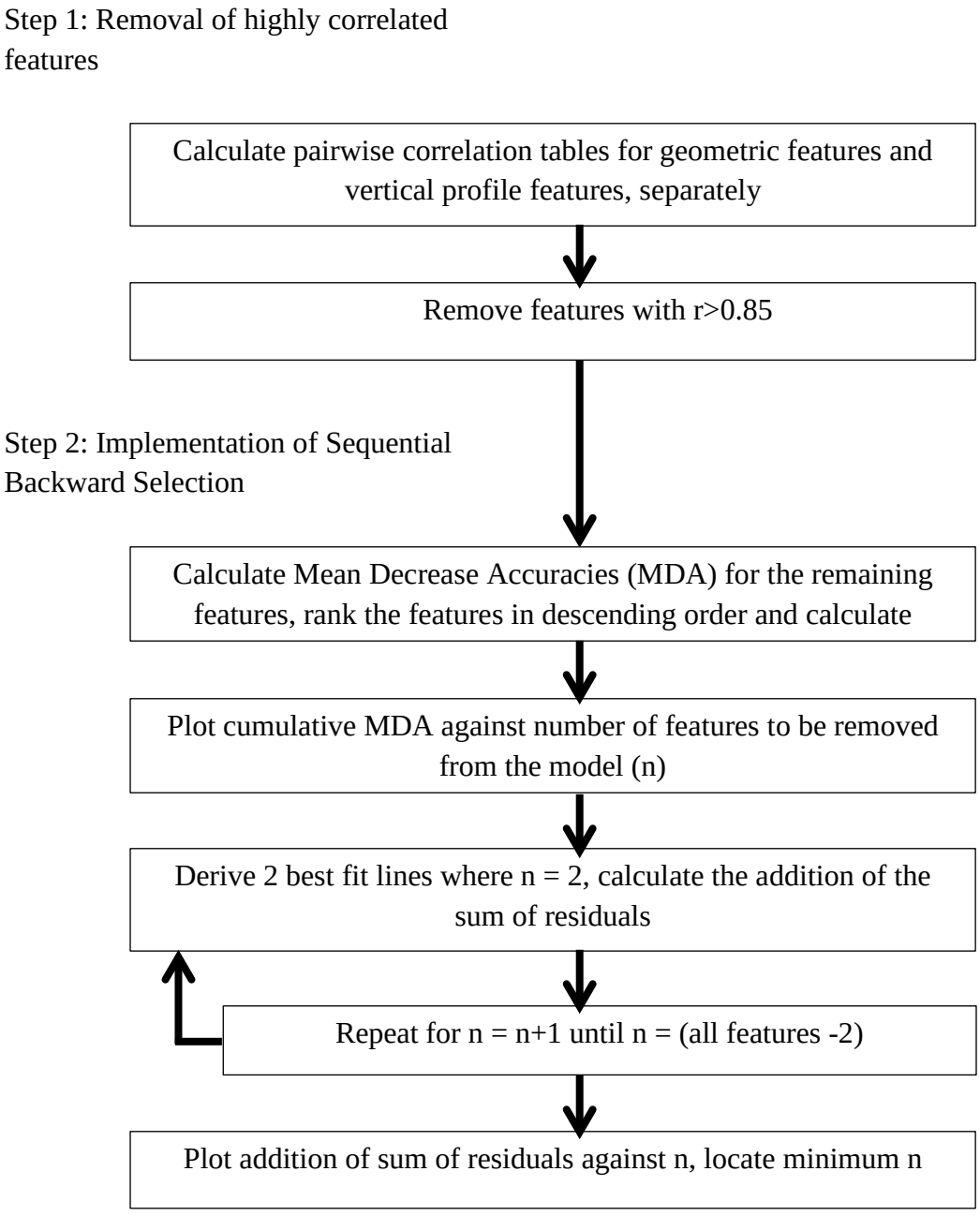

Figure 4-2 Flow chart showing the framework for the feature reduction analysis performed to the geometric and vertical profile classifier.

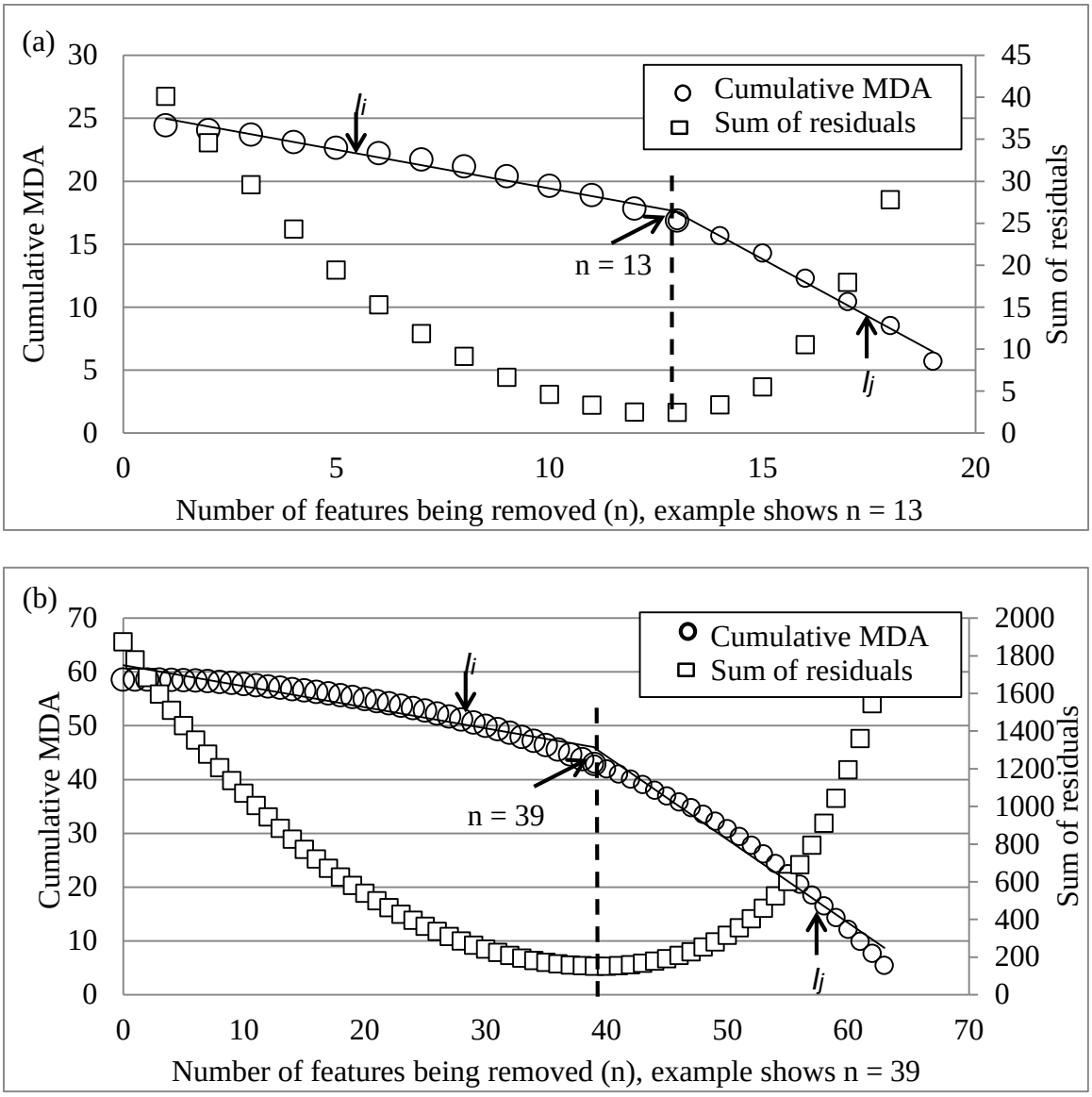

Figure 4-3 (a) Cumulative MDA values for geometric classifiers ((b) Cumulative MDA values for vertical profile classifiers) and residual sum of square residual for fitting two linear lines through the cumulative MDA curve. Dotted line shows where the residual sum of square minimizes. Solid lines represents l_i and l_j at optimized J.

In Figure 4-3(a) the result demonstrates that the best subset of features should be chosen when 13 features are removed (6 remaining) for geometric classifier. In Figure 4-3(b), the result demonstrates that the best subset of feature should be chosen when 39 features are removed (26 remaining) for vertical profile classifier. The advantage of using MDA instead of accuracy for setting up J is that MDA calculation can be obtained from training data alone.

4.2.3 Geometric Classifier, Vertical Profile Classifier and Ensemble methods

Random Forests classifications are performed by using the geometric features and vertical profile features selected from the feature reduction procedure. 25% of the randomly selected (stratified by class size) data is partitioned for the classification scheme and the rest of the data (75%) is partitioned for validation. This 25:75 ratio was chosen from Chapter 3. This process is repeated 20 times with 20 different sets of training and validation samples and an average accuracy is obtained.

In this section an ensemble model is described that has been constructed to combine decisions sequentially and then in parallel. First the classifiers are run individually, from the training data, with the margin of each LiDAR tree recorded from Equation (4.3). From the validation data, the proportion vote for each LiDAR tree is recorded and the pseudo margin for the validation data can be calculated from Equation (4.4). Second, a procedure is adopted to automatically filter out LiDAR trees that are potentially misclassified by the first classifier by examining the *MG* of the training data and *PG* of the validation data. Let σ be the *MG* boundary in the validation data that

separates trees that are potentially incorrect from the trees that are potentially correct. Then, σ is estimated from the training data. From the frequency distribution of the LiDAR trees that are correctly classified for at least 80% (and less than 80%) in Figure 4-4, σ is obtained from testing σ = 0,0.05…1.00. Figure 4-4 (a) shows the frequency distribution of LiDAR trees that are correctly classified for at least 80% (and less than 80%) with 20 randomly selected sample subsets. Since both distributions are equally important hence the frequency count for both distributions are normalized with a maximum of 1. Figure 4-4 (b) shows the normalized frequency distribution of Figure 4-4 (a). By testing with different σ values, Figure 4-4 (c) shows that the total incorrect classification minimizes at σ =0.45 with this given dataset. The *MG* for the incorrect classification at each σ is obtained; σ is chosen such that the total misclassification is minimized (at σ = 0.45). Third, from the first classifier, LiDAR trees with *PG* less than 0.45 are filtered out and run with the other classifier, ultimately accepting the classification decision from the classifier that has a larger PG (parallel ensemble). The ensemble processes are repeated 20 times and mean ensemble classification accuracy is obtained. The scheme of the ensemble method is depicted in Figure 4-5 for the Geometric classifier as the base classifier.

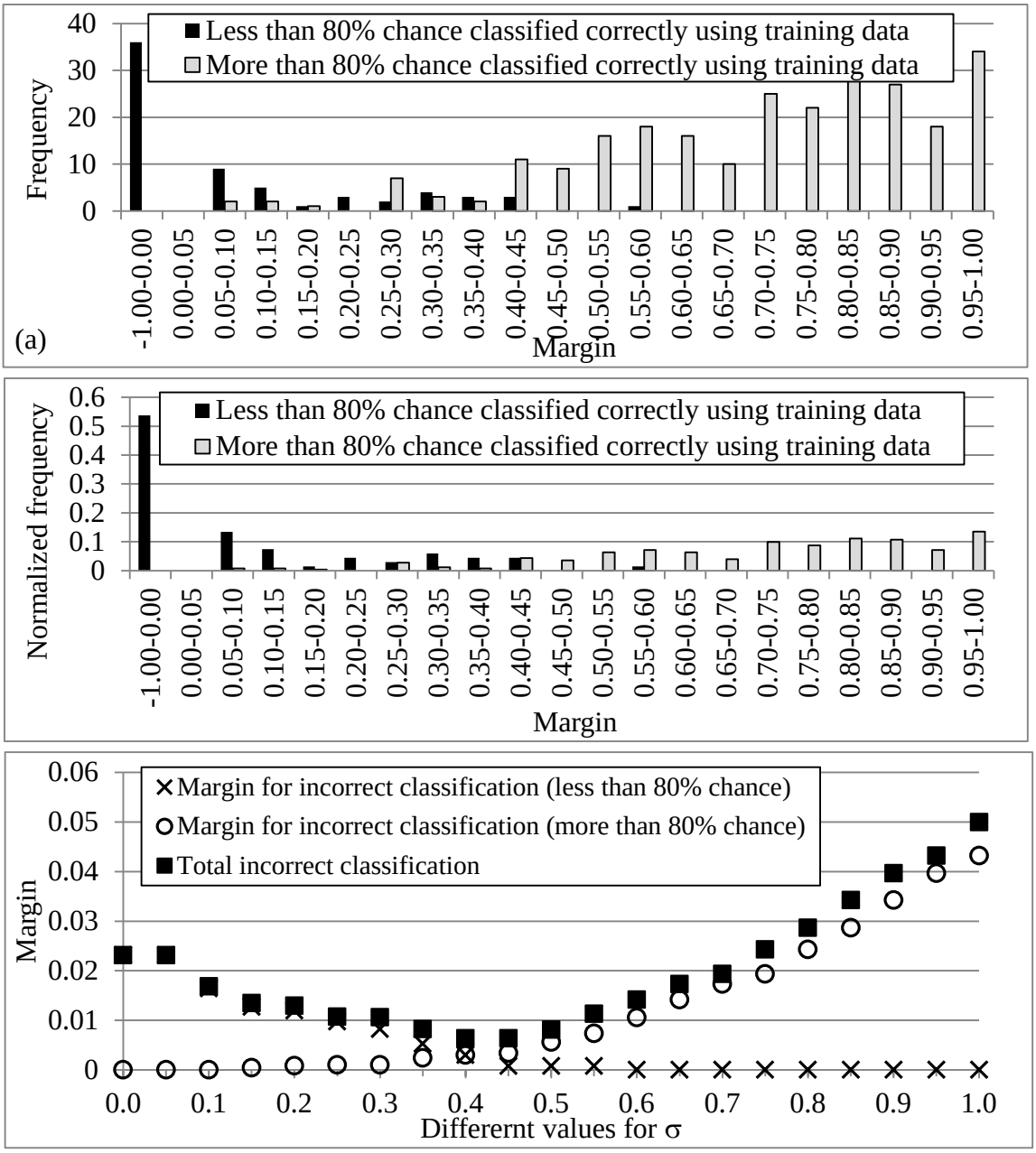

Figure 4-4 (a) Frequency distribution of the LiDAR trees that are correctly classified for at least 80% (and less than 80%) with 20 randomly selected sample subsets. (b) Normalized frequency distribution of (a). (c) Margin for incorrect classification (for more and less than 80% chance) and total margin for incorrect classification at different σ.

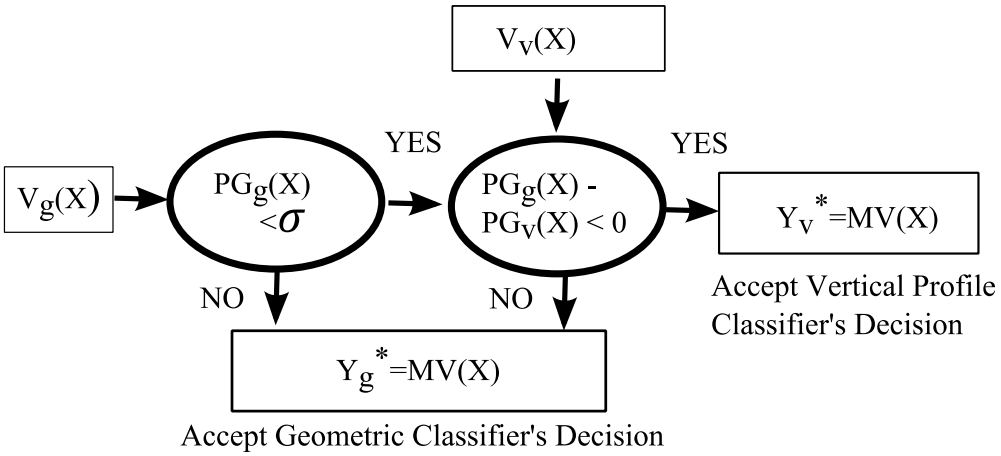

Figure 4-5 Summary of the ensemble method using Geometric classifier as base classifier. MGg, MDv = margin obtained from Geometric classifier and Vertical profile classifier respectively. Voteg, Votev = Vote proportion obtained from Geometric classifier and Vertical profile classifier respectively. Yg*, Yv* = Final prediction from Geometric classifier and Vertical profile classifier respectively.

The densities of individual trees are reduced by removing every other point (with respect to GPS time recorded by the scanner). Therefore, each LiDAR tree will have its original density level (40 pulses / m^2), and reduced to 20 pulses / m^2, 10 pulses / m^2, 5 pulses / m^2, 2.5 pulses / m^2 and 1.25 pulses / m^2. The classification procedures (individual classifiers separately and jointly) are repeated for the different density levels, Figure 4-6 shows an example of a pine tree with various pulse densities.

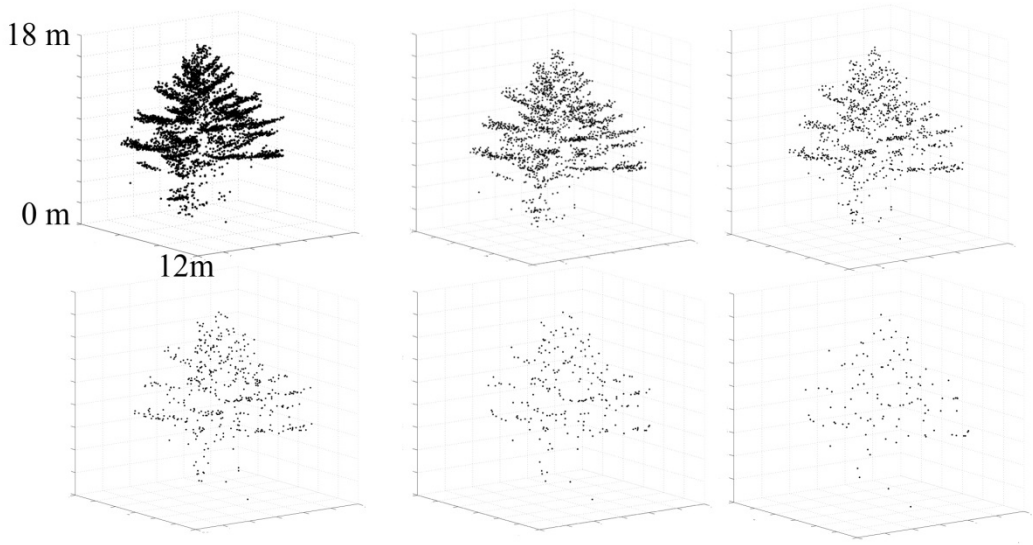

Figure 4-6 Example pine tree with (a) 40 pulses per m^2; (b) 20 pulses per m^2; (c) 10 pulses per m^2; (d) 5 pulses per m^2; (e) 2.5 pulses per m^2; (f) 1.25 pulses per m^2 , source: Ko et al. (2012b)

4.3 Results and Discussion

4.3.1 Selected Feature Tables

From the feature reduction experiment using the geometric classifier, the numbers of features have been reduced from 24 to 6 and 78 to 26 for the vertical profile classifier. The list of selected geometric classifier features are shown in Table 4-2 and the list of selected vertical profile classifier features are shown in Table 4-3.

Table 4-2 List of Selected geometric features

No.	Description
F1	Average derived best bit line segment lengths divided by tree height
F2	Average line segment lengths multiplied by the ratio between tree crown height and tree height
F3	Volume of the tree crown convex hull divided by the number of points in the crown
F4	Average distance from each point to the closest facet of the convex hull
F5	Buffer each LiDAR point outward at a radius of 2% of the tree height, calculate the overlapped volume of the spheres divided by the number of points in the tree crown
F6	Tree crown height divided by the tree height

Table 4-3 List of selected vertical profile featureS, F= first; S= single; L= last; SD= standard deviation; CV = coefficient of variation

10^{th} percentile	50^{th} Percentile	90^{th} percentile
% of canopy return ($V1_{S,}$ $V2_{L)}$		
% return count ($V3_{F}$, $V4_{S,}$ $V5_{L)}$		% return count($V6_{F}$, $V7_{S,}$ $V8_{L)}$
Mean height of canopy return ($V9_{F,}$ $V10_{L)}$		
		SD of height ($V11_{F}$, $V12_{S}$)
SD height for canopy return ($V13_{F}$, $V14_{L)}$		
CV height for canopy return ($V15_{F,}$ $V16_{S}$)		
Kurtosis of variation height for canopy return ($V17_{S,}$ $V18_{L)}$		
Skewness of variation height for canopy return ($V19_{S,}$ $V20_{L)}$		
Mean intensity ($V21_{L)}$		Mean intensity ($V22_{F}$, $V23_{S}$)
SD of intensity ($V24_{F}$)		
CV intensity of canopy return ($V25_{L)}$		
Skewness of variation intensity of canopy return ($V26_{S}$)		

4.3.2 Classification performance

By using the geometric and vertical profile classifiers alone, each method achieved comparable classification accuracy (88.0% and 88.8% respectively). The confusion matrix for classification by geometric and vertical profile derived features is shown in Table 4-4. By using the ensemble method, the accuracy is improved to 91.2% when the geometric classifier is used as the base classifier and to 90.3% when the vertical

profile classifier is used as the base classifier. The confusion matrix for the ensemble results is shown in Table 4-5. By using 75% of the 160 LiDAR trees for classification, repeated 20 times, results in 2400 trees are being assessed.

Table 4-4 Confusion matrix for individual classifier, Bold number: geometric classifier (average accuracy of 88.0%); italic numbers: vertical profile classifier (average accuracy of 88.8%).

		Expected						User's Accuracy (%)	
		Pine		Poplar		Maple			
Predicted	Pine	**856**	*906*	**115**	*132*	**19**	*12*	**86.5**	*86.3*
	Poplar	**123**	*87*	**771**	*736*	**2**	*7*	**86.0**	*88.7*
	Maple	**27**	*13*	**1**	*19*	**486**	*488*	**94.6**	*93.8*
Producer's Accuracy (%)		**85.1**	*90.1*	**86.9**	*83.0*	**95.9**	*96.3*		

When individual classifiers are compared (Table 4-4), both classifiers have the largest errors when trying to separate pine from poplar; this is attributed to the similarity between the vertical point distributions for pine and poplar, with points located mostly at the top of the tree crown. For the geometric classifier, the ratio between the tree crown height and tree height for both genera are also similar, again resulting in confusion. Conversely, highest accuracy is observed in maple classification by both classifiers.

By comparing the results from the geometric classifier alone with the ensemble

classification using the geometric classifier as base classifier and results from vertical profile classifier alone with the ensemble classification using the vertical profile classifier as a base classifier, accuracies for all genera are improved (except for producer's accuracy for maple; vertical profile classifier). This implies that using the margin is effective in automatically filtering out LiDAR trees that are difficult to classify by the base classifier. The improvements in accuracies also suggest the ensemble method outperforms the single classifier alone.

Table 4-5 Confusion matrix for ensemble classification, Bold number: geometric classifier as base classifier (average accuracy of 91.2%); italic numbers: vertical profile classifier as base classifier (average accuracy of 90.3%).

		Expected						User's Accuracy (%)	
		Pine		Poplar		Maple			
Predicted	Pine	**903**	*908*	**94**	*109*	**5**	*21*	**90.1**	*87.5*
	Poplar	**86**	*80*	**786**	*777*	**3**	*5*	**89.8**	*90.1*
	Maple	**17**	*18*	**7**	*1*	**499**	*481*	**95.4**	*96.2*
Producer's Accuracy (%)		**89.8**	*90.3*	**88.6**	*87.6*	**98.4**	*94.9*		

4.3.3 Density Results

The relationship between classification accuracy with different pulse densities is shown in Figure 4-7, with error bars showing the standard error for each point within the

20 trials. For all classifiers, the accuracy stays at about the same level for 40, 20 and 10 pulses·m^{-2}, and then begins decreasing at 5, 2.5 and 1.25 pulses·m^{-2}. The geometric classifier and vertical classifier show similar results for all density levels except at 1.25 pulses·m^{-2}, where the accuracy for vertical profile classifier is 3% lower than geometric classifier. Also, ensemble classification shows higher classification accuracy at all density levels. It can be concluded that there are opportunities to reduce the pulse density to 10 or 5 pulses·m^{-2}, resulting in lower costs in data acquisition and processing handling for comparable results, although the tradeoff between classification accuracy and pulse density will need to be considered if the pulse densities decrease further.

4.4 Conclusions

In this Chapter applied ensemble methods is performed to combine features derived from the geometry of LiDAR points reflected from individual trees with features derived from vertical point distribution. The advantage of using geometric features is that these feature tie close relationship to the biophysical interpretation of trees and the advantage of vertical profile features is that they are computationally simple. Table 4-4 shows that individual classifiers make different decisions, for example the producer's accuracy for pine has the largest difference. The differences indicate there are potentials for improving accuracy after combining the classifiers. By combining the decisions made by the two classifiers, the classification accuracy improved from 88.0% to 91.2% if geometric classifier is being used as base classifier and 88.8% to 90.3% if vertical profile classifier is being used as base classifier. Since the original accuracies (with single classifier alone)

are already very high, the marginal improvement that has been made is very challenging. The ensemble method has improved the overall classification accuracy however the contribution of ensemble to a single classifier has not been studied in this chapter.

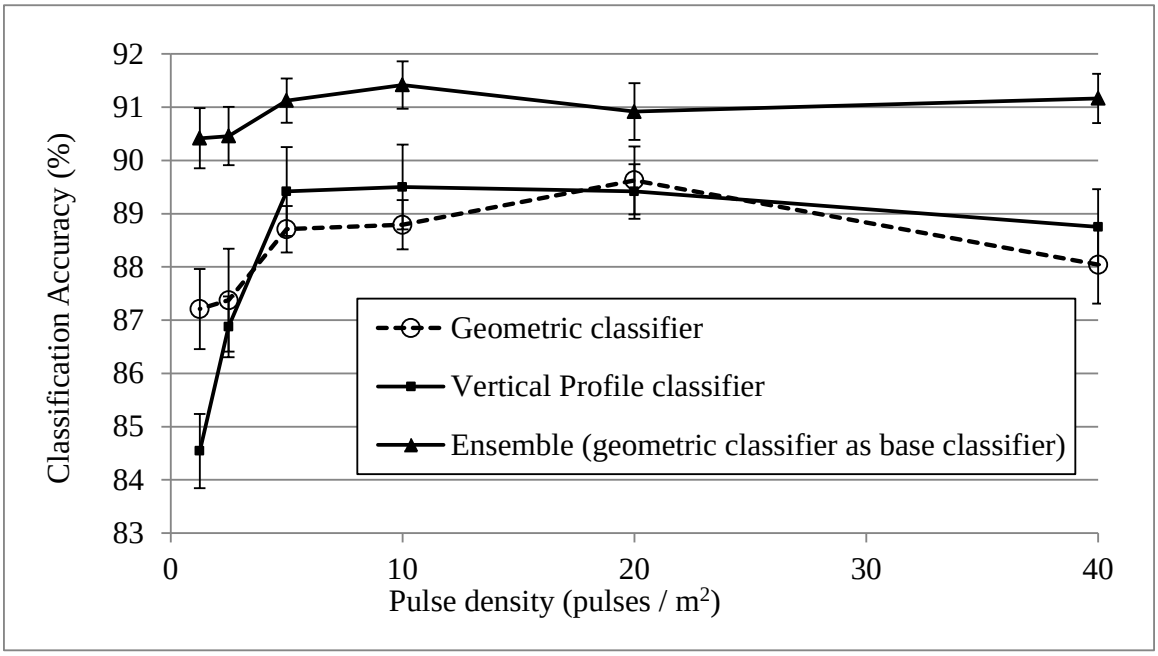

Figure 4-7 Classification accuracy of Geometric classifier, Vertical Profile classifier and Ensemble Classifier, using Geometric classifier as base classifier at different LiDAR pulse density levels.

In order to study the performance of the classifiers with the reduction of LiDAR point density, an experiment studying the relationship between classification accuracy and point density is performed. Result shows that (Figure 4-7) similar classification accuracy can be obtained at lower density level (down to 5 pulses /m^2) and accuracy

dropped for individual classifiers as well as ensemble classifier when density level is lower than 5 pulses / m^2. Indicating that there is a trade-off between classification accuracy and pulse density, however, the trade-off is higher when the pulse density falls beyond 5 pulses / m^2. This result shows there is room for reducing the pulse density to trade for slightly lower classification accuracy, and this normally results in lower costs.

Chapter 5 Classification with Unknown Class

The detection of unknown class in a classification system is essential and often a common problem. This situation arises when the number of classes in the training sample is less than the number of classes in the population. For supervised classifications, training samples are required and the selection (or method of selection) of training sample (training class size, sample quality) is often a subjective choice. This chapter will use ensemble methods to classify tree genera with airborne LiDAR (Light Detection and Ranging) data, including an "unknown" class that does not exist in the training sample. This chapter also shows that by altering the training sample composition, classification accuracy result varies. This indicates classification accuracy can be improved by selecting training samples with certain criteria rather than random sampling. This chapter uses Random Forests as base classifiers with six derived geometric features from the LiDAR data. The training sample contains three tree genera (pine, poplar and maple) and the validation sample contains four labels (pine, poplar, maple and "unknown"). Classification accuracy improved from 72.8% if training samples are selected randomly (with stratified sample size), to 93.8% if samples are selected with criteria; and from 88.4% to 93.8% if ensemble method is being used.

5.1 Introduction

The use of LiDAR (Light Detection and Ranging) data for tree species / genera classification has proven successful in research such as Holmgren and Persson (2004); Holmgren et al. (2008); Brandtberg (2007); Kato et al. (2009); Ørka et al. (2007, 2009); Vauhkonen (2008, 2009, 2010); Korpela et al. (2010) and Kim et al. (2009, 2011) and is gaining more attention because of its ability to obtain 3D information. However, there ar`e two common challenges for using LiDAR to classify tree species / genera and have not been discussed in the literature. The first challenge is the existence of tree species (classes) in the forest that has not been field validated. In supervised classification, this is the problem when the validation data has more classes than the training data. In this case, normally an "unknown class" will be assigned to the extra class(es), e.g. Mantero et al. (2005) and Muñoz- Marí et al. (2007).

The second challenge is the acquisition of consistent 3D tree object, due to scanning angle, often portions of the tree are occluded by another object (especially trees). Moreover, when trees are growing in a closely together, branches (and therefore LiDAR points) are often mingled together and become problematic when one try perform single tree segmentation. Trees appeared to have multiple tree tops could also be confused with multiple single trees. Furthermore, in different environmental conditions and age of trees, same tree species can appear very differently in the acquired LiDAR data. This results a large variation in training and validation samples, in terms of appearance as well as quality. However, this paper take advantage of the variability observed from data collected within the same genera. Training data set that contain the

most variability were able to educate the base classifiers with better decision boundaries and hence improve the overall classification accuracy.

As discussed in Chapter 2, there are three main approaches for decomposing a multi-class classification problem into a series of binary classification problems. The OVO, OVA and ECOC, this chapter will use OVA for decomposition, where each binary classifier is trained on solving only one genus. The two advantages of using OVA for this problem is that first, this ensemble method is flexible to concept change. In the future, if another genus is being added to the dataset, one can add another base classifier without changing the rest of the classifier system. Secondly, this classification system using OVA classifiers has the capability of classifying trees that does not belong to any of the training labels, generating an "unknown" class. This achieves one of the goals for this research and results also show that this method outperforms the original multi-class classifier.

It is not always easy to obtain a complete, single segmented tree for data analysis with LiDAR data and the quality of segmented tree can vary within the dataset. As a result, the frequency distribution of segmented LiDAR tree versus data quality is not uniform. Consequently, when training data is randomly selected from the population, the problem of imbalance training (with respect to data quality) arises. Normally, the problem of imbalance data distribution refers to the differences in the number of samples for each class (Chawla et al. 2004, Fernández et al. 2013) and is a common problem in real world data simply because some phenomena occurred infrequently naturally. Classifiers generated from imbalanced training sample will be specialized in classifying

the majority classes and therefore bias the results towards the majority class. At the data level solution, there are three common techniques to overcome this problem 1) oversampling the minority class 2) under sampling the majority class or 3) a combination of both methods (Chawla et al. 2004, Fernández et al. 2013, Barandela et al. 2004, Weiss 2004, Japkowicz 2000). The goal of these strategies is to diversify the sampling distribution in terms of the number of samples per class and these studies show that the distribution of training sample is important in affecting the classification accuracy. However, this chapter investigates this problem with a slightly different perspective. On top of diversifying the number of sampling distribution, this chapter also diversifies the data quality in the training samples. RandomForests is use as the base binary classifiers and the RandomForests algorithm solve the problem of diversifying number of training sample by bootstrap aggregating (bagging) in two ways. First, the same number of samples are drawn from the minority and majority class; balancing the minority / majority sample (Chen et al. 2004). Second, a heavier penalty is placed on misclassifying the minority class and gives the class a higher weight (Chen et al. 2004), and the final label takes the majority voting of the individual classification trees. Results generated from RandomForests will be used to quantify training sample quality and this matter will be discussed further in the method section.

5.2 Methods

5.2.1 Overview of the methodology

As discussed in Chapter 2, there are two major ways of combining classifiers, in parallel and in sequence. This chapter has performed classification with the two models (parallel and sequential), but because parallel model shows better classification accuracy, therefore, the discussion will be limited to the parallel model. Also, one of the goals for this chapter is to design a classification scheme that is capable for classifying the negative samples (the non-pine, non-poplar and non-maple), therefore OVA decomposition is more suitable for this task. OVA decomposition will be used with parallel model for classifying pine, poplar, maple and "unknown" with Random Forests as base classifier. This section will contain three main components: 1) Selection of training and validation sample, 2) Using Random Forests as base classifiers and 3) Parallel ensemble model. Random Forests classification will be discussed first although it is the second step of overall method because the other components require the understanding of Random Forests.

5.2.2 Random Forests Classification

Random Forests is being used in two different areas in this paper. First is for quantifying the quality of each LiDAR tree and this information can be used for the final selection of training data. The second is for base classifier construction, the three base classifiers are: h_p (classifier that will produce class labels of pine (p) and non-pine (p')));

h_o (classifier that will produce class labels of poplar (o) and non-poplar (o')); and h_m (classifier that will produce class labels of maple (m) and non-maple (m')).

Random Forests algorithm itself is also an ensemble classifier, the final classification labels are predicted by combining multiple classification trees for categorical data, or regression trees for continuous data trained from a subset of the data (Breiman, 2001, Liaw and Wineder, 2002). Random Forests uses approximately 63% of the data for training (in-bag data) and therefore uses approximately 37% of the data (out of bag data) for validation. The main input variables for Random Forests relevant for this chapter are 1) training sample labeled with known genera and a description of the geometric features is listed in Table 5-1, the detail for deriving these features can be found in Ko, et al. (2013). 2) The number of feature variables randomly sampled at each split (mtry, 2 for this research paper), 3) the number of trees generated within each iteration (Ntree, 1000 for this research paper) and 4) Minimum size of terminal node (Nodesize, 1 for this research paper). Random Forests produce the following output relevant for this paper: 1) a classification scheme generated using in-bag training data, 2) the percentage of mis-classification rate using OOB data, and 3) average vote calculated for each class, each LiDAR tree, where final prediction of the OOB data is made by the maximum average vote.

Table 5-1 Description of the six geometric features

No.	Description
F1	Average derived best bit line segment lengths divided by tree height
F2	Average line segment lengths multiplied by the ratio between tree crown height and tree height
F3	Volume of the tree crown convex hull divided by the number of points in the crown
F4	Average distance from each point to the closest facet of the convex hull
F5	Buffer each LiDAR point outward at a radius of 2% of the tree height, calculate the overlapped volume of the spheres divided by the number of points in the tree crown
F6	Tree crown height divided by the tree height

5.2.3 Training sample selection

Training samples are selected from the pine, poplar and maple only whereas the validation samples have more than three genera. This is to mimic the fact that one can only able to field validate certain tree genera but there are more genera than measured in the field. This research originally select training data with stratified random selection, the previous study (Ko, et al. 2013) shows a 25% : 75% ratio for training and validation is optimal for this dataset and therefore this chapter will perform the following experiments with the same ratio. As discussed, this chapter suggests that diversifying data quality in the training sample will improve classification accuracy. In order to validate this hypothesis, first, data quality need to be quantified. To achieve this goal, a sensitivity analysis is performed to study the effect of diversity measurement on classification accuracy.

To quantify the quality of segmented LiDAR tree for training data selection, Random Forests is first run with three class label predictions (pine, poplar and maple). Since only known LiDAR trees are used for training (160 trees), this part of the assessment do not require "unknown" trees and therefore are removed from this assessment. Random Forests randomly select approximately 37% of the data (OOB data) as validation and a prediction is made for each LiDAR tree that are selected as OOB data. With *Ntree* = 1000, 1000 different OOB data sets are selected and predicted, the ratio of how many times a LiDAR tree has been predicted correctly to the number times a LiDAR tree has been selected as an OOB data is calculated. If the ratio is small, it means that the particular LiDAR tree cannot be correctly predicted easily and vice versa. Hence, this ratio is used as an assessment. First, the frequency distribution over the ratio for 10 bins is plotted, from the 160 trees. Then, 47 trees are selected for training (25% of 186 trees) and therefore, with an even distribution, one would keep approximately 5 trees for each bin in order to maximize the diversity of quality in the training sample. However, some of the bins have less than 5 trees in the dataset. As a result, to maximize the quality diversity, for bins that have less than 5 trees, all the trees within that particular bin is selected for training. For bins that have more than 5 trees, the bins are downsized to 5 trees and distribute the excessive trees as evenly as possible among all bins. The original data contains 67 pines, 59 poplars and 34 maples and using this method, 19 pines, 18 poplars and 10 maples are selected for training.

5.2.4 Ensemble architecture

The training samples (47 trees) from the previous section are being used as training the base classifiers and the rest of the trees (139 trees) are used as validation data. The same training data are used for training the three base classifiers, h_p, h_o and h_m with maximized diversity, allowing the base classifiers to learn as much variability as possible. Each of these classifiers are binary classifiers; for h_p, trees will be labeled as *p* (positive) for pines and *p'* (negative) for poplars and maples; for h_o, trees will be labeled as *o* (positive) for poplars and *o'* (negative) for pines and maples and for h_m, trees will be labeled as *m* (positive) for maples and *m'* (negative) for pines and poplars. One of the goals is to be able to correct detect "unknown" class, where three base classifiers vote for negative.

The parallel model is summarized in Table 5-2, where h_p, h_o and h_m are the base classifiers. In the cases where there are no conflict in decision between the base classifiers, (case 1, 2, 3, 8), the final decision will be made by the classifier voted for a positive case (case 1, 2 and 3). For case 8, where all three classifiers vote for negative (case 8), the tree will be labeled as "unknown". In the cases where classifiers are conflicting each other (case 4 to case 7), the final decision will be made by the classifier that has a larger maximum vote calculated for its positive label. Table 5-2 Summary of the parallel ensemble model

Case #	Decision made by h_p	Decision made by h_o	Decision made by h_m	Final Decision
1	p	o’	m’	p
2	p’	o	m’	o
3	p’	o’	m	m
4	p	o	m’	If max vote (p)>max vote(o), p; else, o
5	p’	o	m	If max vote(o)>max vote(m), o; else, m
6	p	o’	m	If max vote(p)>max vote(m), p; else, m
7	p	o	m	Label with max(max vote)
8	P’	o’	m’	u

p= “pine”; p’= "non-pine”; o= “poplar”; p’= "non-poplar”; m= “maple”; m’= "non-maple”; u= “unknown”

5.3 Results and discussion

5.3.1 Quality diversity and training sample selection

To quantify the quality of the LiDAR trees data, the ratio between the number of times a tree can be classified correctly and number of times it has been selected as OOB data for prediction is calculated. Figure 5-1 (a) shows the frequency distribution of the ratio over 10 bins for 160 trees and Figure 5-1(b) shows the frequency distribution of the ratio for selected training samples (47 trees). In Figure 5-1 (a), it shows that most of the LiDAR tree data collected are considered easily identifiable, with a ratio between 0.8 - 1.0, meaning these trees have above 80% chance to be classified correctly. To visualize the differences between the calculated ratios, Figure 5-2(a) shows six examples of LiDAR trees with lower ratio values and Figure 5-2(b) shows six examples with higher ratio values. From Figure 5-2(b), it can be observed that these LiDAR trees have less

occlusion, overlapping tree crown points with neighbourhood trees and larger in size (more data points per tree) and are therefore easier to be classified.

When training samples are selected randomly, the distribution of the training data will be similar to the distribution shown in Figure 5-1 (a), with several bins being empty because of the low frequency count of the original data. If the base classifiers are trained base on randomly selected samples, more emphasis will be placed on the LiDAR trees that has high ratio. Hence, this chapter maximizes the diversity of the training sample by including as much variability as possible in terms of data quality, Figure 5-1 (b) shows the frequency distribution the training sample selection.

To verify the relationship between the diversity of training dataset and classification accuracy, trees from 0.0-0.1 bin are removed manually (9 bins left from Figure 5-1(b)) and the same number of removed trees are redrawn and redistributed to the other bins in order to maintain the same total number of training samples. The classification accuracy is obtained from this new distribution and this procedure is repeated for 0-0.2 bin (8 bins left) until only 1 bin is left. Figure 5-3 shows the result of the analysis performed with and without the “unknown” samples in the validation data, and is performed by using regular Random Forests (without ensemble method) and parallel ensemble for comparison.

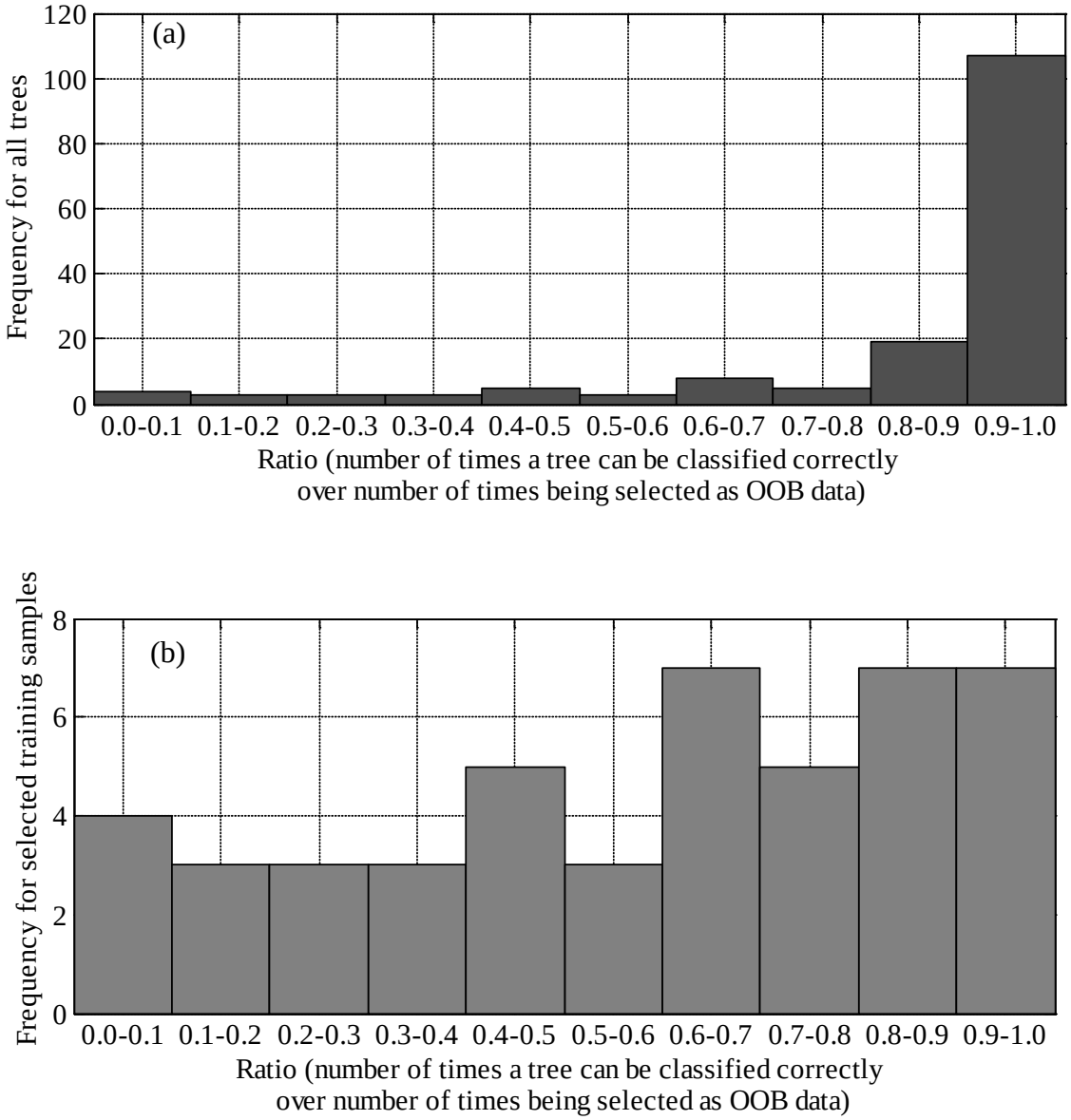

Figure 5-1 (a) Frequency distribution of ratio (number of trees can be classified correctly over number of times being selected as OOB data) for all trees (160 trees); (b) Frequency distribution of ratio for selected training sample (47 trees)

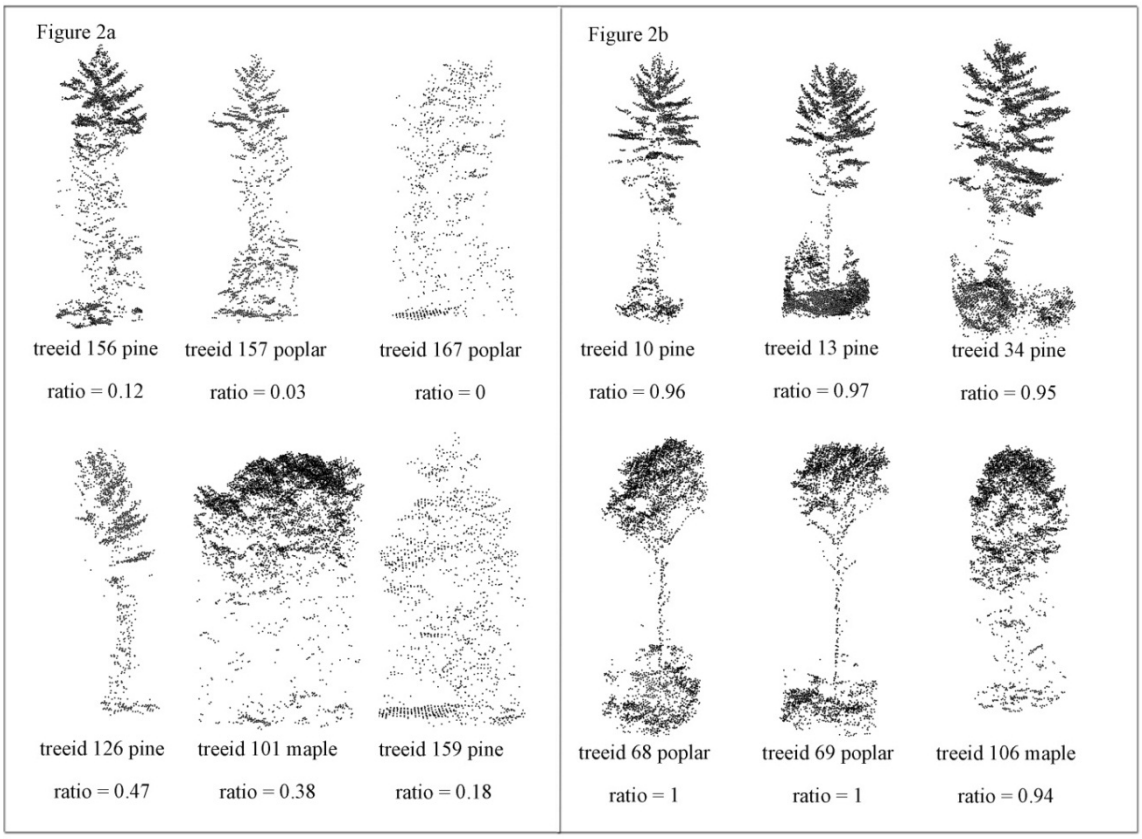

Figure 5-2 (a) example LidAR trees that has a lower (less than 0.4) ratio (number of times it can be classified correctly over number of times it has been selected for prediction); (b) example LidAR trees that has a higher (larger than 0.9) ratio

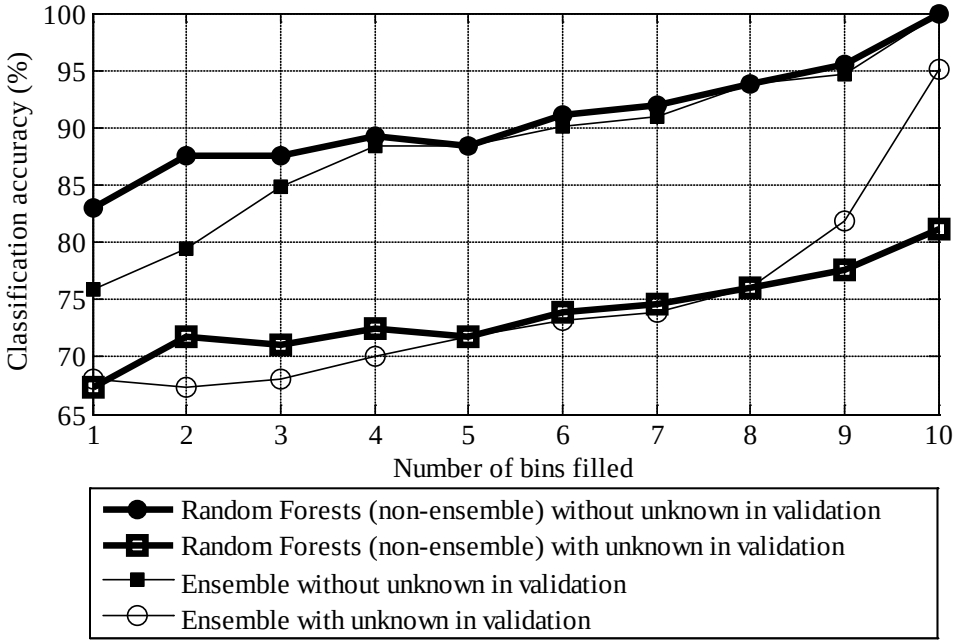

Figure 5-3 Classification accuracy changes over number of bins being filled

In Figure 5-3, result shows that as the number of bins being filled increases (more diversified training data), the classification accuracies increases, with or without unknown samples in the validation data. Also, when unknown samples are included in the validation data, ensemble classification outperformed regular Random Forests by 4.3% and 13.9% when the 9^{th} and 10^{th} bin are (0.0-0.1 bin and 0.1-0.2 bin) being filled, respectively. Meaning that these samples, although have lower quality, or their feature distributions deviate from the majority of the training sample, are important in classifying the "unknown" data in the validation sample. Also, when "unknown" data are removed from the validation sample, ensemble classification shows lower classification accuracy,

by 8.0% and 7.1% when classified with only 1 or 2 bins (0.9-1.0 bin and 0.8-1.0 bins) and ensemble methods should be avoided in this case.

5.3.2 Classification for Random Forests and ensemble Random Forests

This section performs two comparisons, first is to compare the classification results obtained from random sampling (ensemble classification) versus the diversified training sample (ensemble classification) (Figure 5-1(b)). The second is to compare the classification results obtained from Random Forests and ensemble Random Forests, both using the diversified training sample. Table 5-3 shows the confusion matrix obtained from ensemble classification using random sampling and Table 5-4 shows the confusion matrix obtained from ensemble classification using diversified training sample. The overall accuracy improved from 72.8% to 93.8% (Table 5-3 and Table 5-4) when the training samples have maximum diversification. The commission and omission error has reduced 60% and 53% respectively for “unknown” class when the training sample is diversified. This is because “unknown” classes are labeled based on the classification accuracies of negative samples of the base classifiers.

Table 5-3 Confusion matrix obtained from ensemble classification using random sampling, values are averaged over 20 different random samples

		Measured				
		pine	*poplar*	*maple*	*unknown*	Commission error
predicted	*pine*	37.45	3.85	0.05	6.75	0.22
	poplar	4.05	35.7	0.2	2.45	0.16
	maple	1.15	0.15	23.9	12.55	0.37
	unknown	4.15	1.8	0.55	4.25	0.60
Omission error		0.20	0.14	0.03	0.84	
Overall Accuracy: 72.8%						

Table 5-4 Confusion matrix obtained from ensemble classification using diversified sampling, values are averaged over running ensemble 20 times

		Measured				
		pine	*poplar*	*maple*	*unknown*	Commission error
predicted	*pine*	47.00	0.00	0.20	3.40	0.07
	poplar	0.00	41.00	0.00	2.95	0.07
	maple	0.00	0.00	23.80	2.00	0.08
	unknown	0.00	0.00	0.00	18.65	0.00
Omission error		0.00	0.00	0.01	0.31	
Overall Accuracy: 93.8%						

When the training samples have low diversity, it indicates most of the training samples appear similarly and therefore the distributions of the features related to the specific class are narrow. As a result, any validation data that deviates from the training

data will be classified as negative (high omission error, 84% in Table 5-3, reduced to 31% in Table 5-4). Conversely, when the training sample contains larger variety, the distributions of values for the features that train the base classifiers are broader allowing a broader definition of a specific class, reduce the chance of classifying negative sample as positive (commission error of “unknown” class reduced from 60% to 0%), . The error of the “unknown” class also come from the difficulties in differentiate genera that appear similarly under LiDAR reflected points, it is found that birch and oak trees are normally mistaken as maple trees; spruces and larch trees are normally mistaken as pine trees. Figure 5-4 shows an example of each mistaken pair.

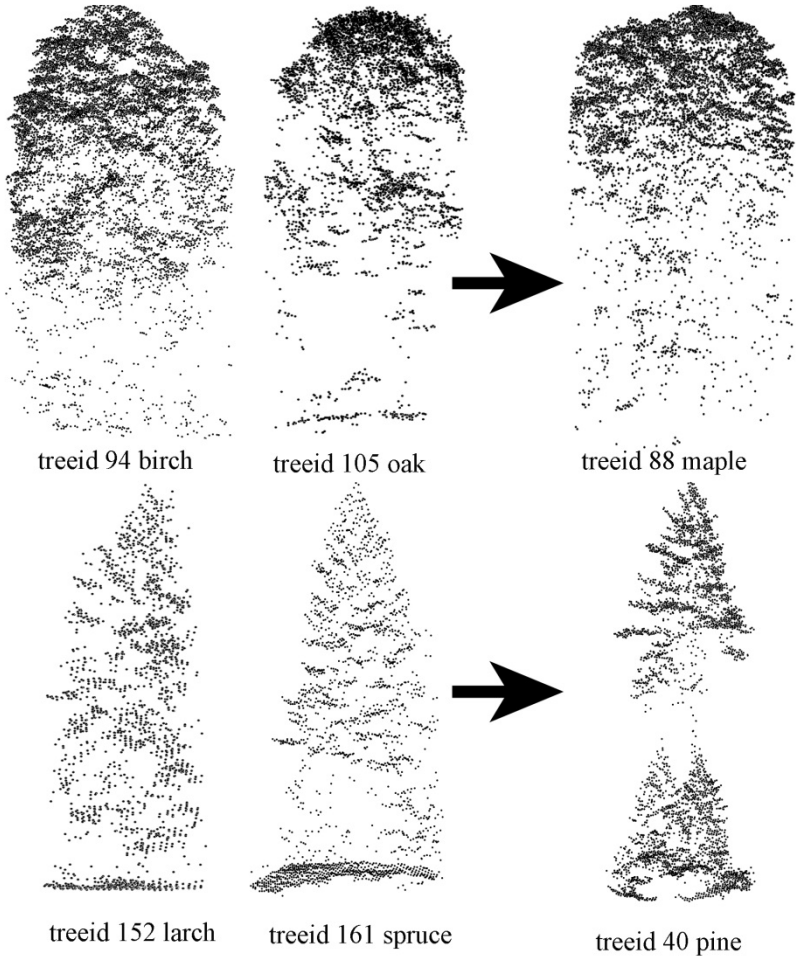

Figure 5-4 Example misclassification from birch and oak to maple; larch and spruce to pine

The second comparison is made between the classification accuracy of Random Forests with classification accuracy of ensemble methods. Since the training sample contains only pine, poplar and maple, by default, the class labels generated from RandomForests will be the three mentioned classes. However, in order to be able to compare with ensemble methods, additional condition is included so that “unknown”

classes can be identified with Random Forests alone. For each validation sample in RandomForests, the final class label is assigned by the class with the maximum vote. The condition is that if the maximum vote calculated for the particulate tree is less than 67%, it should be classified as “unknown” instead of one of the three classes. 67% is obtained from dividing 100% into three classes, assuming if more than 67% of the votes are voting for that particular class, the prediction is valid. Hence, the trees are labeled with “unknown” if the vote is less that 67%. Table 5-5 shows the confusion matrix obtained from Random Forests using diversified samples, values are averaged over running ensemble 20 times. By comparing Table 5-5 with Table 5-4, ensemble Random Forests classification yields an overall accuracy of 93.8% whereas Random Forests alone yield an overall accuracy of 88.4%. The omission error for Random Forests alone is less than ensemble methods but the commission error for classifying poplar for Random Forests is higher. With ensemble methods, the omission error for pine, poplar and maple is also lower.

Table 5-5 Confusion matrix obtained from Random Forests classification using diversified sampling, values are averaged over running ensemble 20 times

		Measured				
		pine	*poplar*	*maple*	*unknown*	Commission error
Predicted	*pine*	39.60	0.00	0.00	0.00	0.00
	poplar	7.40	25.00	2.55	4.20	0.36
	maple	0.00	0.00	38.45	1.00	0.03
	unknown	0.00	1.00	0.00	19.80	0.05
Omission error		0.16	0.04	0.06	0.21	
Overall Accuracy: 88.4%						

5.4 Conclusions

This chapter shows the importance of diversifying the quality of training samples in terms of improving classification accuracy. The quality of LiDAR tree samples can vary due to reasons such as shadow from scanning angle, overlapping to the neighborhood trees or trees that appear to have multiple tree tops making tree crown isolation difficult. This chapter hypothesizes that when the classifiers are trained with samples that contain as much variability as possible, it will have a better decision boundary between the positive and negative class labels. The nature of the data collected for this research contains majority of the samples that have similar properties (Figure 5-1(a)), by randomly selecting the training data, the training data will have similar distribution as the population. In that case, the majority of the training data will be similar to each other, any validation data that deviates from the training data will be mis-classified. To solve this problem, the quality of collected samples are quantified by the OOB results generated

from Random Forests and training samples are re-drawn with maximum quality diversity. With a broader definition of each genus, results shows that classification accuracy can be improved from 72.8% to 93.8%, both using ensemble methods, the accuracy especially improved for the classification of negative labels.

The base classifiers are the decomposition of *k* binary classification models for *k* classes. By combining these base classifiers instead of using Random Forests alone as a multi-class classifier, a better accuracy result can be obtained (improved from 88.4% to 93.8%). There are several reasons why this is a better method than the traditional multi-class classification. 1) Ensemble methods are able to generate "unknown" class without pre-defined threshold, 2) the method does not require the presence of "unknown" in the training data. 3) The implementation of binary classification is simple, in the future, if another class label is being collected in the field, an additional binary classifier will be built but all the previous training results can be re-use without any changes. 4) The method of combination can be altered and additional rules can be implemented to improve the aggregation of information such as Table 5-2. Although the classification accuracy can be improved by implementing the ensemble of binary classifiers, the omission error for the "unknown" class is still the highest (Table 5-4). This is because of the confusion of the similarities shared between genera such as Birch and oak trees are normally mistaken as maple trees; spruces and larch trees are normally mistaken as pine trees (Figure 5-4). This problem can be resolved if new features for classification can be developed that are independent from geometry (the six features described in Chapter 3). In conclusion, this chapter shows that by diversifying the training sample characteristic,

classification performance can be improved. Also, the ensemble of binary classifiers allows us to incorporate the “unknown” class without much extra effort and also improves overall classification accuracy.

Chapter 6 Conclusions

As the collection of LiDAR data has becoming economically available, the use of LiDAR data for environmental applications becomes popular. Using LiDAR data for forestry applications can automate measurements for forest structure and the aim is to replace the traditional labor-intensive, time consuming manual measurements. Tree species / genera is one of the most important information among all attributes but it is not easy to obtain directly from only the coordinate of the point clouds.

As discussed in Chapter 2, there are mainly three successful approaches on obtaining tree species / genera information. First approach relies on information located near or at the top of the tree, second approach relies on derivation of vertical point profile information. The third approach relies on the derivation of geometric features, mostly related to the outer shape of the tree, such as convex hull and alpha shape features. The geometric features derived in this research not only considers the outer shape of individual tree, the geometric features derived in Chapter 3 also considered the structures and features observed inside the tree crown. Six of the 24 geometric features are selected for classifying pine, poplar and maple in the study area, using Random Forests as the classifier. Instead of using the default OOB error for evaluation, where about 63% (in bag) of the data are used for training the classifier, leaving 37% of the data for validation. A sensitivity analysis was performed to study the relationship between classification accuracies and percentage of data used for training. Results shows that by using 25% of the data for training is sufficient for obtaining a classification accuracy of 88.3%. In

conclusion, Chapter 3 provides a fundamental research by building features that are related to the geometry of the tree, which is different from the existing approaches. Three main conclusions can be drawn from this Chapter. 1) Geometric features tie close relationship to the biophysical interpretation of trees and can be used for genera classification. 2) Geometric features provide an alternate perspective for classification. 3) The variability in classification accuracies in different field sites suggests the quality of LiDAR tree data affects classification accuracy and this issue will be further addressed in Chapter 5. This research relies on single tree data that are already segmented from the LiDAR scene, which is a manual process. An automatic tree delineation method has to be implemented for operational applications.

Chapter 4 further advance the ideas developed from Chapter 3 by combining the classification features derived from Chapter 3 and the conventional features, vertical profile features through an ensemble classification. Also, a feature reduction method (Sequential Backward Selection) that utilizes the results produced from Random Forests (MDA – mean decrease accuracy) is fully discussed in this chapter. This process reduces the number of features for geometric classifier (24 to 6) and vertical profile classifier (78 to 26). Feature reduction is an important process for any classification model, there is usually a trade-off between classification accuracies and the number of features being used for the model. An optimum number of features should be selected for the model that balances model complexity and accuracy. The goal of any effective ensemble classification system is to increase classification accuracy, this chapter combines two classifiers first sequentially and then in parallel. Without using a predefined threshold,

trees that are potentially problematic for the first classifier are automatically filtered out. Final decision is made by using both classifiers for those problematic LiDAR trees. At 40 pulses / m^2, the average classification accuracy for vertical profile classifier and geometric classifier is 88.0% and 88.8% respectively for 20 random training samples. The overall classification accuracy improves to 91.2% by the ensemble system. In this chapter, an additional sensitivity analysis is performed to study the relationship between point density and classification accuracy. The classification accuracies for the individual classifiers as well as ensemble system are expected to drop as point density decreases. This analysis is performed to study the lower limit of point density for this type of study, the result shows that classification accuracies does not drop significantly until point density dropped to 1.25 pulses / m^2 from the original 40 pulses / m^2. This indicates there is a potential for lowering point density for classification with a slight tradeoff for classification accuracy. The two main contribution of this chapter are 1) implemented an automatic feature reduction method using Random Forests. 2) Study the potential of using ensemble classification to combine geometric and vertical profile features.

Chapter 5 has two major contributions, the first is by introducing an “unknown” class in the validation class label, this increases the practicality of the classification model. The classification of environmental objects normally involve obtaining training data from field work, therefore the classifier will be trained with field collected data. However it is not unusual to have more classes in the validation data than the training data. For example, one normally would not survey every tree in the forest and therefore cannot guarantee training class label fully represents the validation class labels. Hence,

the inclusion of “unknown” class allows categorizing validation samples that does not belong to any one of the training labels into a reasonable label instead of forcing it as one of the training labels. This is done by breaking up a multi class problem into a series of binary classification problems. By combining these binary classifiers, not only an “unknown” class can be included, the classification accuracy also increased from 88.4% (by using Random Forests alone) to 93.8% (combination of binary classifiers). The second contribution of Chapter 5 involves the discussion and quantification of LiDAR tree sample quality. As discussed in Chapter 3, some LiDAR trees segmented from the LiDAR scene suffers from problem of occlusion and high overlapping tree crown from neighborhood vegetation (lower quality samples), lowering the overall classification accuracy. Chapter 5 quantify this qualitative description, quality, by calculating the number of times a particular LiDAR tree sample is being correctly classified in the OOB data divided by the number of times a sample is being selected as OOB data. A sensitivity analysis is performed to study the relationship between classification accuracy and the inclusion of these low quality samples in the training data. This is done by diversifying the quality of training data as much as possible. Results shows that although these samples have low classification accuracy itself, by including them in the training data allow the base classifiers to learn about the variability within each genera and improved the overall classification especially the for the “unknown” class. Overall classification accuracy increased from 72.8% (using random sampling) to 93.8% (diversified sampling).

As a whole, this dissertation advances the science of using discrete airborne LiDAR data for tree genera classification. Geometric features are derived for describing the internal and external features of the tree crown, rather than relying on only the vertical point profile of the LiDAR point distributions. Feature reduction method allow the simplification of classification model and therefore increase model efficiency. The ensemble of geometric classifier and vertical profile classifier improves overall classification accuracy. The decomposition of a multi-class classification problem into a series of binary classification problem allows the inclusion of "unknown" class in the validation data. Finally, this dissertation discusses the fact that by diversifying the training data quality improves the overall classification accuracy especially for the "unknown" class.

References

Aldred, A.H. and Bonnor, G.M., 1985. Application of airborne lasers to forest surveys. Information Report PI-X-51. *Canadian Forestry Service, Petawawa National Forestry Institute*, 62 p.

Ali K. M. and Pazzani, M. J., 1996. Error reduction through learning multiple descriptions. *Machine Learning*, 24(3): 173–202.

Baillard, C. and Maître, H., 1999. 3-D reconstruction of urban scenes from aerial stereo imagery: a focusing strategy. *Computer Vision and Image Understanding*, 76(3): 244-58.

Baltsavias, E., 1999. Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3): 199-214.

Barandela, R., Valdovinos, R.M., Sánchez, J.S. and Ferri, F.J., 2004. The Imbalanced Training Sample Problem: Under or Over Sampling? In: Fred, A., Caelli, T.M., Duin, R.P.W., Campilho, A.C., de Ridder, D. (eds.). *SSPR&SPR 2004. LNCS,* 3138: 806-814. Springer, Heidelberg.

Barilotti, A., Crosilla, F. and Sepic. F., 2009. Curvature analysis of LiDAR data for single tree species classification in alpine latitude forests. *Laser scanning*, 1-2 September 2009, Paris, France (IAPRS, Vol. XXXVIII, Part 3/W8).

Becker, S., and Haala, N., 2007. Refinement of building facades by integrated processing of LiDAR and image data. *The International Archives of the*

Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part3/W49A), 7–12.

Bell, D.A., Guan, J.W. and Bi, Y., 2005. On combining classifier mass functions for text categorization, *IEEE Trans. on Knowledge and Data Engineering*, 17(10):1307-1319.

Bortolot, Z.J., 2006. Using tree clusters to derive forest properties from small footprint LiDAR data. *Photogrammetric Engineering & Remote Sensing*, 72(12): 1389-1397.

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.

Brandtbert, T., 2002. Individual tree-based species classification in high spatial resolution aerial images of forests using fuzzy sets. *Fuzzy Sets and Systems*, 132: 371-387.

Breiman, L., 1998. Arcing Classifiers. *Annals of Statistics*, 26: 801-824

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

Brovelli, M., Cannata, M., and Longoni, U., 2003. Lidar data filtering and DTM interpolation within GRASS. *Transactions in GIS*, 8(2): 155–174.

Bryll, R., Gutierrez-Osuna, R. and Quek, F., 2003. Bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition Society*, 36: 1291-1302.

Bucksch, A. and Fleck, S., 2011. Automated detection of branch dimensions in woody skeletons of fruit tree canopies. *Photogrammetric Engineering & Remote Sensing*, 77(3): 229-240.

Buddenbaum , H. Schlerf, M. and Hill, J., 2005. Classification of coniferous tree species and age classes using hyperspectral data and geostatistical methods. *International Journal of Remote Sensing*, 26(24): 5453–5465.

Chawla, N.V., Japkowicz, N. and Kolcz A., 2004. Editorial: Special Issue on Learning from Imbalanced Data Sets. *ACM SIGKDD Explorations Newsletter*, 6(1): 1-6.

Chen, C., Liaw, A. and Breiman, L., 2004. Using random forest to learn imbalanced data. *Technical Report 666*, University of California, Berkeley.

Clark, M. L., Roberts, D. A. and Clark, D. B., 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, 96(3-4): 375-398.

Cochrane, M. A., 2000. Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensing*, 21(10): 2075-2087.

Coulston, J. W., Moisen, G. G., Wilson, B. T., Finco, M. V., Cohen, W. B. and Brewer, C.K., 2012. Modeling percent tree canopy cover: a pilot study. *Photogrammetric Engineering and Remote Sensing*, 78(7): 715-727.

Côté, J. F., Fournier, R.A., and Egli, R., 2011. An architectural model of trees to estimate forest structural attributes using terrestrial LiDAR. *Environmental Modelling and Software,* 26(8): 761-777.

Côté, J. F., Widlowski, J.L., Fournier, R.A. and Verstraete, M.M., 2009. The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial LiDAR. *Remote Sensing of Environment,* 113(5): 1067-1081.

Dietterich, T. G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40(2): 139-158.

Dietterich, T. G. and Bakiri, G., 1995. Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research,* 2: 263–286.

Duan, K., Jagath, J. C., Rajapakse, Nguyen, M.N., 2007. One-Versus-One and One-Versus-All Multiclass SVM-RFE for Gene Selection in Cancer Classification. *EvoBIO,* 47-56.

Duda, R., Hart. P. and Stork, D., 2000. Pattern Classification (2nd Edition). *Wiley-Interscience*, Hoboken, NJ, 654 p.

Eckert, K., 2004. Proper tree and vegetation management makes major differences. *Natural Gas Electricity*, 21(5): 1-8.

Falkowski, M. J., Evans, J.S., Martinuzzi, S., Gessler, P.E. and Hudak, A.T., 2009. Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. *Remote Sensing of Environment*, 113(5): 946 – 956.

Fernández A., López V., Galar M., José del Jesús M. and Herrera F., 2013. Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches. *Knowledge-Based Systems*, 42: 97-110.

Galar M., Fernández A., Barrenechea E., Bustince H. and Herrera F., 2011. An Overview of Ensemble Methods for Binary Classifiers in Multi-class Problems: Experimental Study on One-vs-One and One-vs-All Schemes. *Pattern Recognition*, 44(8): 1761-1776.

Gillis, M.D. and Leckie, D.G., 1996. Forest inventory update in Canada. *The Forestry Chronicle*, 72: 138-156.

Gillis, M.D., 2001. Canada's national forest inventory (responding to current information needs). *Environmental Monitoring and Assessment*, 67: 121- 29.

Hashemi, S., Yang, Y., Mirzamomen, Z. and Kangavari, M., 2009. Adapted One- versus-All Decision Trees for Data Stream Classification. IEEE Transactions on Knowledge and Data Engineering, 21(5): 624-637.

Hastie T. and Tibshirani R., 1998. Classification by pairwise coupling. *Annals of Statistics*, 26 (2): 451–471.

Hilker, T., Wulder, M.A. and Coops, N.C., 2008. Update of forest inventory data with lidar and high spatial resolution satellite imagery. *Canadian Journal of Remote Sensing*, 34(1): 5-12.

Hill, R.A., and Thomson, A.G., 2005. Mapping woodland species composition and structure using airborne spectral and LiDAR data. *International Journal of Remote Sensing*, 26(17): 3763-3779.

Hollaus, M., 2006. Large scale applications of airborne laser scanning for a complex mountaineous environment. Dr. Ing. thesis, *Vienna University of Technology*, 111 p.

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.

Hornberg, A., 2006. Handbook of Machine Vision, Wiley, 821 p.

Hsu, C.W. and Lin, C.J., 2002. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13(2): 415-425.

Hudak, A. T., Crookston, N.L., Evans, J.S., Hall, D.E. and Falkowski, M.J., 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment*, 112(5): 2232 – 2245.

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): 1339–1366.

Irish, J. L. and Lillycrop, W. J., 1999. Scanning laser mapping of the coastal zone: The SHOALS system. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54:123–129.

Ituen, I., and Sohn, G., 2010. The Way Forward: Advances in Maintaining Right- Of-Way of Transmission Lines. *GEOMATICA*, 64(40): 451-462.

Ituen, I., Sohn, G. and Jenkins, A., 2008. A Case Study: Workflow Analysis of Power Line Systems for Risk Management, *IAPRS* 37, Part B3b: 331-336.

Jain, A. and Zongker, D., 1997. Feature selection: evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine,* 19(2): 153-158.

Japkowicz, N., 2000. Learning from Imbalanced Data Sets: a Comparison of Various Strategies. *AAAI Workshop on Learning from Imbalanced Data Sets*, AAAI Press, Menlo Park, CA.

Jwa, Y., and Sohn, G., 2012. A piecewise catenary curve model growing for 3D power line reconstruction. *Photogrammetric Engineering & Remote Sensing*, 78(11): 1227-1240.

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*, 113(6): 1148-1162.

Kavzoglu and Colkesen, 2013. An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping. *International Journal of Remote Sensing*, 34(12): 4224-4241.

Kern, W. F. and Bland, J. R., 1948. Solid Mensuration with Proofs, 2nd ed. NewYork: Wiley, 172p.

Kim, S., Hinckley, T. and Briggs D., 2011. Classifying individual tree genera using stepwise cluster analysis based on height and intensity metrics derived from airborne laser scanner data. Remote Sensing of Environment, 115(12): 3329 – 3342.

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., 2009. Classification for deciduous and coniferous trees using airborne LiDAR and internal structure reconstructions. *Proceedings of SilviLaser,* 14-16 October, Texas.

Ko, C., Sohn, G. and Remmel, T. K., 2012a. A comparative study using geometric and vertical profile features derived from airborne LiDAR for classifying tree genera. *ISPRS Annals of Photogrammetry*, Remote Sensing and Spatial Information Sciences, I-3, pp. 129-134.

Ko, C., Sohn, G. and Remmel, T. K., 2012b. The impact of LiDAR point density on classifying tree genus: using geometric features and vertical profile features. SilviLaser 2012, September 16-19, Vancouver, British Columbia, Canada.

Ko, C., Sohn, G. and Remmel, T. K., 2012c. Mapping tree genera using discrete LiDAR and geometric tree metrics. *BOSQUE* , 33(3): 313-319.

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.

Koerich, A., Sabourin, R. and Suen, C., 2005. Recognition and verification of un-constrained handwritten words. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10): 1509-1522.

Korpela, I., Dahlin, B., Schäfer, H., Bruun, E., Haapaniemi, F., Honkasalo, J., Ilvesniemi, S., Kuutti, V., Linkosalmi, M., Mustonen, J., Salo, M., Suomi, O. and Virtanen, H., 2007. Single-tree forest inventory using LiDAR and aerial images for 3D treetop positioning, species recognition, height and crown width estimation. *IAPRS* Volume XXXVI, Part 3 / W52, 2007. pp. 227-233.

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 Fennica,* 44(2): 319–339.

Korpela, I., Tokola, T., Ørka, H.O. and Koskinen, M., 2009. Small footprint discrete-return LiDAR in tree species recognition. *Proceedings of the ISPRS*, 2-5 June 2009, Hannover, Germany (Workshop).

Koukoulas, S., and Blackburn, G.A., 2005. Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LIDAR and multi-spectral remotely sensed data. *International Journal of Remote Sensing*, 26(3): 431-455.

Kraus, K. and Pfeifer, N., 2001. Advanced DTM generation from LIDAR data. *In International Archives of Photogrammetry and Remote Sensing*, Annapolis, MD, XXXIV (3/W4): 3–30

Kumar, S., Ghosh, J. and Crawford, M. M., 2002. Hierarchical Fusion of Multiple Classifiers for Hyperspectral Data Analysis. *Pattern Analysis & Applications, Special Issue on Fusion of Multiple Classifiers,* 5(2): 210-220.

Leckie, D. and Gillis, M., 1995. Forest inventory in Canada with an emphasis on map production. *The Forestry Chronicle*, 71: 74-88.

Lefsky, M.A., Cohen, W.B., Parker, G.G. and Harding, D.J., 2002. LiDAR remote sensing for ecosystem studies. *BioScience*, 52(1):19-30.

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

Lin, C., Thomson, G., Lo, C.S. and Yang, M.S., 2011. A multi-level morphological active contour algorithm for delineating tree crowns in mountainous forest. *Photogrammetric Engineering & Remote Sensing*, 77(3): 241-249.

Linares, R., Mateo, G. and Castro, A., 2003. On combining classifiers for speaker authentication. *Pattern Recognition*, 36(2): 347-359.

Liu, Y. and Zheng, Y.F., 2005. One-against-all multi-class svm classification using reliability measures. *In Neural Networks, 2005. IJCNN '05. Proceedings*. 2005 IEEE International Joint Conference on, 2(2): 849 - 854.

Lodha, S. K., Kreps, E. J., Helmbold, D. P. and Fitzpatrick, D., 2006. Aerial LiDAR Data Classification Using Support Vector Machines (SVM). *Proceedings IEEE Symposium on 3D Data Processing, Visualization, and Transmission*, pp. 567-574.

Magnusson, M., Fransson, J. E. S. and Holmgren, J., 2007. Effects on estimation accuracy of forest variables using different pulse density of laser data. *Forest Science*, 53(6): 619-626.

Mallet, C. and Bretar, F., 2009. Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(1): 1-16.

Mallet, C., Bretar, F., Soergel, U., 2008. Analysis of Full-Waveform Lidar Data for Classification of Urban areas. *Photogrammetrie - Fernerkundung - Geoinformation* (PFG), 5/2008, pp. 337-349.

Martinuzzi, S., Vierling, L.A., Gould, W.A., Falkowski, M.J., Evans, J.S., Hudak, A.T. and Vierling, K.T., 2009. Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sensing of Environment*, 113(12): 2533 – 2546.

Means, J., Acker, S., Harding, D., Blair, J., Lefsky, M., Cohen, W., Harmon, M.and McKee, A., 1999. Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western cascades of Oregon. *Remote Sensing of Environment*, 67(3): 298–308.

Milgram, J., Cheriet, M. and Sabourin, R., 2006. "OneAgainst One" or "One Against All": Which One is Better for Handwriting Recognition with SVMs?" In G. Lorette, editor, *Proceedings of 10th International Workshop on Frontiers in Handwriting Recognition*. Suvisoft, 2006.

Mitchell, J., Glenn, N.F., Sankey, T., Derryberry, D.R., Hruska, R. and Anderson, M.D., 2011. Small-footprint lidar estimations of sagebrush canopy characteristics. *Photogrammetric Engineering & Remote Sensing*, 77(5): 521-530.

Morrison, R., Low, B., Kucera, G. and Kucera H., 1999. A spatial data warehouse for NFIS pp. 242-247 in *Proceedings of the 13th Annual Conference on Geographic Information Systems (GIS 99*), March 1-4, 1999, Vancouver, BC. Canadian Forest Service, Pacific Forestry Center, Victoria, BC.

Muñoz-Marí, J., Bruzzone, L. and Camps-Valls, G., 2007. A support vector domain description approach to supervised classification of remote sensing images. *IEEE Transactions on Geoscience and RemoteSensing*, 45(8):2683–2692.

Na, X., Zhang, S., Li, X., Yu, H. and Liu, C., 2010. Improved Land Cover Mapping using Random Forests Combined with Landsat Thematic Mapper Imagery and Ancillary Geographic Data. *Photogrammetric Engineering & Remote Sensing*, 76(7): 833-840.

Næsset, E., 2004. Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sensing of Environment*,91(2): 243–255.

Nelson, R., Krabill,W. and Maclean, G., 1984. Determining forest canopy characteristics using airborne laser data. *Remote Sensing of Environment*, 15(3): 201–212.

Niccolai, A., Niccolai, M. and Oliver, C.D., 2010. Point set topology extraction for branch and crown-level species classification. *Photogrammetric Engineering & Remote Sensing*, 76(3): 319-330.

Ørka H. O., Næsset, E. and Bollandsås, O. M., 2007. Utilizing airborne laser intensity for tree species classification. *International Archives of Photogrammetry & Remote Sensing, Remote Sensing and Spatial Information Sciences,* Part 3/W52. 36, pp. 300–304.

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 Envornment, 113(6): 1163-1174.

Oza, N.C., Polikar, R., Kittler, J. and Roli F. (eds), 2005. Multiple ClassiØers Systems. *Proceedings of the 6th International Workshop on MCS*, LNCS 3541, Springer-Verlag GmbH.

Park, H., Lim, S., Trinder, J., and Turner, R., 2010. 3D surface reconstruction of terrestrial laser scanner data for forestry. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 25 - 30th July 2010, 4366- 4369.

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

Persson, Å., Holmgren, J., Söderman, U. and Olsson H., 2004. Tree species classification of individual trees in Sweden by combining high resolution laser data with high resolution near infrared digital images. International Conference NATSCAN 'Laser- Scanners for Forest and Landscape Assessment – Instruments, Processing Methods and Applications', 3–6 October 2004, Freiburg, Germany. *In*

International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVI, 8/W2, pp. 204–207.

Petrie, G. and Toth, C.K., 2009. Topographic Laser Ranging and Scanning: Principles and Processing. (Shan, J. and Toth, C.K. Ed.). CRC Press, Boca Raton. 590p.

Pfeifer, N., and Winterhalder, D., 2004. Modelling of tree cross sections from terrestrial laser scanning data with free-form curves. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36: 76 - 81.

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/.

Rahman, M. and Gort, B., 2008. Individual Tree Detection Based On Densities of High Points of High Resolution Airborne LiDAR. *In proceedings GEOBIA, 2008* - Pixels, Objects, Intelligence: GEOgraphic Object Based Image Analysis for the 21st Century, University of Calgary, Calgary, Alberta, Canada, pp. 350-355.

Rahman, M., Gorte, B.G.H. and Bucksch, A., 2009. A new method for individual tree delineation and undergrowth removal from high resolution airborne LiDAR Laserscanning. *Proceedings ISPRS Workshop Laserscanning 2009*, September 1-2, France, IAPRS, XXXVIII (3/W8), 2009, Volume XXXVIII, Paris, France, pp. 283-288.

Reitberger, J., Heurich, M., Krzystek, P. and Stilla, U., 2007. Single tree detection in forest areas with high-density LiDAR data. *International Archives of Photogrammetry*, Remote Sensing and Spatial Information Sciences, 36: 139-144.

Rifkin and Klautau, 2004. In Defense of One-Vs-All Classification. *Journal of Machine Learning Research*, 5: 101-141.

Rottensteiner, F., Trinder, J., Clode, S. and Kubik, K., 2005. Using the Dempster- Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion*, 6(4): 283-300.

Ruta, D. and Gabrys, B., 2000. An Overview of Classifier Fusion Methods. *Computing and Information Systems*, 7(1): 1-10.

Samadzadegan, F., Bigdeli, B. and Ramzi, P., 2010. A multiple classifier system for classification of lidar remote sensing data using multi-class svm. *in IAPR Multiple Classifier Systems*, Cairo, Egypt, 5997: 254–263.

Schenk, T. and Cshato, B. 2007. Fusing imagery and 3D point clouds for reconstructing visible surfaces of urban scenes. *Proceedings of Urban Remote Sensing*, 2007, Paris.

Schwing, A., Zach, C., Zheng, Y. and Pollefeys, M., 2011. Adaptive Random Forest - How many "experts" to ask before making a decision?", *in Proc. CVPR'11—IEEE International Conference on Computer Vision and Pattern Recognition*, Colorado Springs, United States, 2011, pp. 1337- 1384.

Serpico, S.B., D'Incà, M., Melgani, F. and Moser, G. 2002. A comparison of feature reduction techniques for classification of hyperspectral remote-sensing data. *in Proceedings. SPIE—Conference*. Image and Signal Processing for Remote Sensing VIII, Crete, Greece, pp. 347–358.

Solberg, S., Næsset, E. and Bollandsås, O.M., 2006. Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. *Photogrammetric Engineering & Remote Sensing*, 72: 1369-1378.

Strobl, C., Malley, J. and Tutz, G., 2009. An Introduction to Recursive Partitioning: Rational, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. *Physiological Methods*, 14: 323-348.

Suratno, A., Seielstad, C. and Queen, L., 2009. Tree species identification in mixed coniferous forest using airborne laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6): 683-693.

U.S.-Canada Power System Outage Task Force, 2004. Final report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations, Kennedy Research Laboratories, Arlington, VA, USA, pp. 20, 194, 197.

Vauhkonen, J., Korpela, I., Maltamo, M. and Tokola, T., 2010. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment*, 114(6): 1263-1276.

Vauhkonen, J., Tokola, T., Maltamo, M. and Packalén, P., 2008. Effects of pulse density on predicting characteristics of individual trees of Scandinavian commercial species using alpha shape metrics based on ALS data. Canadian Journal of Remote Sensing, 34(Suppl. 2) S441-S459.

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.

Wagner, W., Ullrich, A., Ducic, V., Melzer, T., Studnicka, N., 2006. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitizing airborne laser scanner. *ISPRS Journal of Photogrammetry & Remote Sensing,* 60 (2): 100-112.

Wehr, A. and Lohr, U., 1999. Airborne laser scanning – an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3): 68-82.

Weiss, G., 2004. Mining with rarity: A unifying framework. *SIGKDD Explorations*, 6(1): 7-19.

Woods, M., Pitt, D., Penner, M., Lim, K., Nesbitt, D., Etheridge, D. and Treitz, P., 2011. Operational implementation of a LiDAR inventory in Boreal Ontario. *Forestry Chronicle*, 87(4): 512- 528.

Wulder, M. A. and Seemann D., 2003. Forest inventory height update through the integration of lidar data with segmented Landsat imagery. *Canadian Journal of Remote Sensing*, 29(5): 536-543.

Wulder, M. A., White, J.C., Bater, C.W., Coops, N.C., Hopkinson, C. and Gang, C., 2012. Lidar plots—A new large - area data collection option: Context, concepts, and case study. Canadian Journal of Remote Sensing, 38(5): 600-618.

Xu, H., Gossett, N., and Chen, B., 2007. Knowledge and heuristic based modelling of laser-scanned trees. *ACM Transactions on Graphics*, 26(4):19, 303-308.

Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M. and Viitala, R., 2011. Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1): 28-37.

Zhou, Z. H., 2012. Ensemble Methods: Foundations and Algorithms. Chapman & Hall / CRC, Boca Raton, FL, 236 pp.