A Random Forest Based Method for Urban Land Cover Classification using LiDAR Data and Aerial Imagery

by

Jiao Jin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Geography

Waterloo, Ontario, Canada, 2012

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Urban land cover classification has always been crucial due to its ability to link many elements of human and physical environments. Timely, accurate, and detailed knowledge of the urban land cover information derived from remote sensing data is increasingly required among a wide variety of communities. This surge of interest has been predominately driven by the recent innovations in data, technologies, and theories in urban remote sensing. The development of light detection and ranging (LiDAR) systems, especially incorporated with high-resolution camera component, has shown great potential for urban classification. However, the performance of traditional and widely used classification methods is limited in this context, due to image interpretation complexity. On the other hand, random forests (RF), a newly developed machine learning algorithm, is receiving considerable attention in the field of image classification and pattern recognition. Several studies have shown the advantages of RF in land cover classification. However, few have focused on urban areas by fusion of LiDAR data and aerial images.

The performance of the RF based feature selection and classification methods for urban areas was explored and compared to other popular feature selection approach and classifiers. Evaluation was based on several criteria: classification accuracy, impact of different training sample size, and computational speed. LiDAR data and aerial imagery with 0.5-m resolution were used to classify four land categories in the study area located in the City of Niagara Falls (ON, Canada). The results clearly demonstrate that the use of RF improved the classification performance in terms of accuracy and speed. Support vector machines (SVM) based and RF based classifiers showed similar accuracies. However, RF based classifiers were much quicker than SVM based methods. Based on the results from this work, it can be concluded that the RF based method holds great potential for recent and future urban land cover classification problem with LiDAR data and aerial images.

Acknowledgements

First, I would like to thank my supervisor, Prof. Jonathan Li, for all the precious comments and insightful advice he has provided me throughout my master research. Without him, this thesis would not have been completed or written. I would also like to express great thanks to Prof. Richard Kelly, Prof. Jane Law, and Prof. David Clausi for agreeing to be part of my thesis committee and reading my thesis.

I am especially grateful to my graduate fellow, Haiyan Guan for her kind assistance with my LiDAR research. I also wish to thank Ms. Susie Castela and Ms. Lynn Finch for assisting me in many different ways.

Lastly, and most importantly, I wish to thank my family for giving me all unconditional support and love during this long journey.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

This chapter introduces and establishes the field of this study. Section 1.1 presents the motivations of this study by summarizing relevant previous methods in the field of urban land cover classification and identifying a research gap in random forests (RF) based classification using LiDAR data and aerial imagery. Section 1.2 states the objectives of this study to fill the research gap. Last, the organization of this thesis is given in Section 1.3.

1.1 Motivations

Urban land cover information has always been crucial due to its ability to link many elements of human and physical environments. Timely, accurate, and detailed knowledge of the urban land cover information derived from remotely sensed data is increasingly required among a wide variety of communities, such as urban and regional planners (Mittelbach and Schneider, 2005; Santana, 2007; Bhatta, 2010), urban morphology scientists (Lo, 2007; Batty, 2008; Schneider and Woodcock, 2008), environmental scientists (Stefanov and Netzband, 2005; Hepinstall et al., 2008), and global change researchers (Small, 2005; Turner et al., 2007; Grimm et al., 2008).This surge of interest has been predominantly driven by the recent innovations in data, technologies, and theories in urban remote sensing. (Weng and Quattrochi, 2007; Yang, 2011b; Zhong and Zhang, 2012).

Traditionally, land cover maps are derived by field surveys with GPS (global positioning system) receivers or by manual human interpretation on hard-copy maps or aerial photographs. Both methods are expensive and time-consuming. In addition, urban object extraction from image data faces three major challenges. First, perspective occlusions may result in insufficient information. Second, weak image features may be extracted because the building boundaries have low contrast. The third challenge is that, if an image feature does not correspond to an object feature, ambiguity in the reconstruction procedure can result (Chen et al., 2008).

The concentration of LiDAR (light detection and ranging) technology has been a substantial increase over recent years because of the rising popularity and availability of the LiDAR data and systems, which obtain dense point measurements using three-dimensional coordinates more directly than traditional surveying and mapping systems, e.g., photogrammetric systems (Shan and Sampath, 2005; Mongus and Zalik, 2012). The principle behind LiDAR remote sensing is the measurement of the distance from the sensor to a reflecting surface, based on the TOF (time-of-flight) measurement of an emitted laser pulse (Wehr and Lohr, 1999). Since the laser pulse can be distended while traveling through the air and reflected by roof, tree, ground, etc., multiple returns of a pulse with different elevation information can be generated. LiDAR systems also record the energy of the returned pulse - the intensity, which is related to the reflective properties of the targets. In this thesis, the term “LiDAR” prefers to airborne discrete-return LiDAR technology, other types of LiDAR (e.g., full waveform LiDAR) will not be discussed here.

During the last decade, airborne LiDAR has become a well-established technology for urban land cover classification due to: (1) rapid and cost-effective acquisition of topographic information at high spatial resolution over complex surfaces where in situ measurements are challenging to acquire (Goodwin et al., 2009), (2) fewer requirements for data pre-processing (Meng et al., 2009), and (3) no negative effects caused by weather conditions, such as cast shadow (Rottensteiner and Clode, 2009). Therefore, LiDAR data has been widely used in a lot of research on urban object classification and extraction (Clode et al., 2007; Dorninger and Pfeifer, 2008; Lee et al., 2008; Weng, 2011).

Considering the great capability of fusing dense laser scanned points with other data sources, various data fusion methods have been used for urban feature extraction, such as high-resolution aerial or satellite imagery (Huber et al., 2003; Sohn and Dowman, 2007), two-dimensional geographic information system (2D GIS) data (Schwalbe, 2005), and ground plans (Vosselman, 2001). The availability of voluminous and high dimensional multisource data, especially in the complex urban environment, poses challenge to information extraction and classification (Sithole and Vosselman, 2003). A dedicated method is required for effective object classification. Due to the complex structure of urban areas, an urban object of a given land cover class (e.g., a single rooftop) may include pixels with heterogeneous reflectance values at a fine scale (Kontoes et al., 2000). However, the performance of traditional and widely-used statistical classification methods is limited in this context (Thomas et al., 2003; Chen et al., 2004; Khoshelham et al., 2010). On the other hand, the state of the art machine learning algorithms,

such as support vector machines (SVM) and RF, are receiving considerable attention in the field of image classification and pattern recognition.

Given the Hughes phenomenon (Hughes, 1968) which refers that an increasing number of features may decrease the classification accuracy, a larger number of training samples are required for supervised classification of fusion data to retain the accuracy. However, larger training size may increase the computational complexity. It is also not easy to obtain in the urban scenes. To solve this problem, two solutions are suggested. One is to develop an advanced classifier with superior classification ability based on small training sample size in high dimensional space. The other is to decrease the number of features using feature selection methods. RF based method is beneficial to solve this problem because of its ability to function as a superior classifier and also as a feature selection tool (Yu et al., 2011). Several studies have shown the advantages of RF in land cover classification (Benediktsson and Sveinsson, 2004; Chan and Paelinckx, 2008; Waske and Braun, 2009; Stumpf and Kerle, 2011). However, very little research has focused on urban areas by fusion of LiDAR data and aerial imagery.

1.2 Objectives of the Study

The overall goal of this study is to explore machine learning techniques for classifying an urban scene (Niagara Falls, ON) using airborne LiDAR data along with aerial imagery. The main objectives of this study are:

- To investigate the performance of RF based classifier for airborne discrete-return LiDAR data and aerial imagery,
- To evaluate the ability of RF as a feature selection tool for urban land cover classification,
- To compare the classification results obtained using the RF based classifier with those obtained using other well-known classifiers.

An effective and operational scheme is expected to extract desired information from LiDAR data and aerial imagery in an urban environment for use in further applications, such as three-dimensional (3D) city modeling and solar potential assessment of buildings.

1.3 Organization of the Thesis

The thesis is organized into five additional chapters.

Chapter 2 reviews literature on urban land cover classification in the field of remote sensing. It describes a variety of land cover classification methods in urban areas, especially using LiDAR data and aerial imagery. A literature review on feature selection methods for land cover classification is also provided.

Chapter 3 introduces two the state-of-the-art classifiers which are mentioned in previous chapters: SVM based and RF based classifiers. Their theoretical background and development are stated. It provides the advantages and limitations of using each classifier. It also defines the parameter selection criteria to achieve the best possible outcome.

Chapter 4 presents in detail the urban land cover classification schemes including those based on SVM and RF. It first describes the study area and datasets used in this study. Then, all the input feature vectors derived from the datasets are listed followed by the training and reference data collected from the study area. Last, the experimental procedures and evaluation methods for feature selection and classification are explained.

Chapter 5 presents and discusses all the classification results derived from the classification schemes presented in Chapter 4. First, a comparison of two feature selection methods which are based on principal component analysis (PCA) and RF is provided. Furthermore, it provides the quantitative and qualitative assessment of all the classification schemes including those via MLC, SVM, and RF based classifiers with and without feature selection using different training sample sizes. Last, the computation time of the SVM based and RF based classifiers is presented.

Chapter 6 presents the findings and conclusions derived from the results of this study. Further suggestions to overcome the remaining challenges are also provided.

Chapter 2
Land Cover Classification Methods: An Overview

This chapter provides an overview of land cover classification approaches, especially for urban areas by integration of LiDAR data and aerial imagery. Section 2.1 presents different land cover classification methods in the field of remote sensing. Section 2.2 reviews feature selection techniques, as well as the comparison methods of different techniques. Finally, Section 2.3 provides a summary of this chapter.

2.1 Classification Methods in Land Cover Classification

Over the past decades, considerable research has been conducted for improving land cover classification performance in three major areas: (1) development and use of advanced classification methods; (2) use of multiple features of remotely sensed data, including texture features and integration of different sensor data; (3) and reduction of the data redundancy (Lu and Weng, 2007; Lu et al., 2007, Gutiérrez et al., 2010). Therefore, the foci of this chapter are the use of a suitable classifier as well as proper feature extraction and selection approach.

In general, land cover classification approaches can be grouped as supervised and unsupervised, or parametric and nonparametric, or pixel-based and object-based, or hard and soft

classification. This section groups classification methods as object-based, unsupervised pixel-based, supervised pixel-based, and knowledge-based classification. A brief description of each category is provided in the following subsection.

2.1.1 Object-based Classification Techniques

An object-based classification (OBC) consists of two major steps. First, image segmentation merges pixels into objects. In this step, the scale of segmentation determines the occurrence or absence of an object class; and the size of an object influents a classification result. Then classification is conducted based on the objects, not an individual pixel. OBC method has proven to be able to provide good performance for multiple sources of data (Herold et al., 2003; Geneletti and Gorte, 2003; Benz et al., 2004; Gitas et al., 2004; Walter, 2004).

Zhu and Troy (2008) applied an object-based classification approach for analyzing and characterizing the urban landscape structure at the parcel level using high-resolution digital aerial imagery (0.6 m pixel size) and LiDAR data. A three-level hierarchical network was built to classify different classes at different levels. Hussain et al. (2011) performed an object-based one-class-at-a-time land cover classification using 0.5 m resolution GeoEye-1imagery and LiDAR data over the City of Port-au-Prince. They concluded that fusing optical imagery and LiDAR data produced adequate classification results in densely populated urban areas. Meng et al. (2012) proposed an object-oriented approach to detect residential buildings from LiDAR and aerial photographs for advanced urban land-use analysis. Although OBC methods have advantages on

urban land cover classification with multispectral remote sensing data (Blaschke, 2010), image segmentation in the first stage of OBC methods can group pixels from different classes into one class, resulting in a decreased classification accuracy (Wang et al., 2004). In addition, prior knowledge on the study areas is required in OBC methods to iteratively search for the optimal segmentation parameters which are crucial to the final classification results.

2.1.2 Unsupervised Pixel-based Classification Techniques

Unsupervised classification methods have been implemented in several studies on land cover classification (Shah et al., 2004; Jiang et al., 2004; Koltunov and Ben-Dor, 2001; Koltunov and Ben-Dor, 2004). No prior definitions of the classes or training data are used in this type of classifers (Duda and Canty, 2002). All the feature inputs are classified into a number of clusters based on the statistical information intrinsic in remotely sensed data, and then the clusters are labeled and merged into meaningful land cover types by analyst. The most commonly used unsupervised classifier is the iterative self-organizing data analysis (ISODATA) (Hall and Ball, 1965) and K-means clustering algorithm (MacQueen, 1967). An unsupervised clustering method was used for automatically extracting buildings from LiDAR data (Hao et al., 2009). In another study, Bartels and Wei (2010) separated object and ground points in LiDAR data using a threshold-free unsupervised classification algorithm called Skewness Balancing. Based on unsupervised stepwise cluster analysis, Kim et al. (2011) classified individual tree genera using airborne LiDAR data. However, unsupervised classifiers require the selection of proper thresholds for determining different land cover types. Its performance mainly relies on the

analyst or the following classifiers to merge clusters into final land cover types. Moreover, other knowledge about remotely sensed data is difficult to incorporate into unsupervised classifier to further improve the results (Jain et al., 2000).

2.1.3 Supervised Pixel-based Classification Techniques

Adapting machine learning methods from pattern recognition and computer vision is a trend in the field of land cover classification in urban areas (Rottensteiner, 2010; Vatsavai et al., 2011). Machine learning is a branch of artificial intelligence and generally concerned with the design and development of methods that allow computers to optimize their performance at tasks by learning from the experience (Mitchell, 1997). In the field of remote sensing, the major part of machine learning algorithms is aiming for a supervised classification which often used in land cover classifications. It provides a simple classification scheme that correctly and quickly extracts objects from complex landscape by learning from training samples.

Classifiers that use Bayes' theorem are one of the traditional technologies in pattern classification. Bayesian classifiers assume specific probability density functions for each class to compute the a-posteriori probabilities. The MLC approach, which is based on the Bayesian framework, is one of the most popular supervised classifier in the field of remote sensing for land cover classification (Bartels and Wei, 2006; Vastsavai et al., 2011). The a-posteriori probabilities are calculated using the probability density function for each class and the a-priori probabilities which refer to the probability that each class occurs in the dataset. Then a pixel is assigned to the

class with the highest a-posteriori probability based on the maximum a posterior rule (MAP) which is a simplified case of the Bayes rules. MLC is based on a Gaussian distribution model estimated from the training data. In this way, the probability density function for each class can be generated by the mean value and the covariance matrix. However, the performance of the MLC approach could be not very effective and superior, due to the need of more computations per pixel, and poor results accuracy especially when the number of training samples is limited.

Neural networks (NN) have been evaluated in many remote sensing applications for the classification of hyperspectral (Ratle et al., 2010), SAR (Bruzzone et al., 2004), high-resolution image (Del Frate et al., 2007), multispectral (Dixon and Candade, 2008), and multisource datasets (Koetz et al., 2008). Previous studies have shown that their results are better than or at least as equal as conventional classifiers such as MLC and spectral angle mapper (SAM) (Petropoulos et al., 2010). A neural network is, according to Haykin (1999), a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. No a priori assumption of statistical distribution of the data is required in NN. Among different proposed neural network algorithms, the most widely-used approaches for land cover classification are error backpropagation multi-layer perceptrons (MLP) networks (Murthy et al., 2003; Estep et al., 2004; Verbeke et al., 2004; Fuller, 2005), radial basis function (RBF) networks (Foody, 2004a; Keramitsoglou et al., 2005), adaptive resonance theory (ART) networks (Muchoney and Strahler, 2002; Pellizzeri et al., 2003; Liu et al., 2004), and self-organizing maps (SOM) networks (Kurnaz et al., 2005; Filippi and Jensen, 2006). However, the NN classifier may not be practical in some

remote sensing applications, due to its complex computational structure and requirement of large number of reliable training data (Mas and Flores, 2008). Another limitation of NN is that there is no exact solution for optimal parameter selection (Stathakis, 2009). It is difficult to choose the optimal number of hidden layers and hidden nodes for neutral network construction.

Decision trees (DT) have also been used for extracting land cover information from remote sensing data in a variety of research (Hansen et al., 1996; Friedl and Brodley, 1997; DeFries and Chan, 2000; Pal and Mather, 2003; Wang and Li, 2008). In a DT, each node denotes a test on an attribute value, each branch represents an outcome of the test, and tree leaves represent classes or class distributions (Han and Kamber, 2001). In contrast to NN, DT can achieve easily interpretable rules and fast computation process, as well as an embedded ability for feature selection (Wang and Li, 2008). Wu et al. (2009) applied decision tree classifier to discriminate among urban land use types by fusion of LiDAR data and imagery along with relevant GIS datasets. In another study (García-Gutiérrez et al., 2011), a DT approach was employed for automatic land cover/ land use classification using LiDAR data. The results showed that DT classifier obtained better accuracy compare to NN classifier. Hofle et al. (2012) discriminated urban vegetation using full-waveform LiDAR data based on DT technique. For future work, the authors suggested improve the accuracy by using a feature selection prior to the DT classification. However, the decision tree can over-fit the training data. Another limitation is that tree pruning is usually required to remove the least reliable lower branches caused by noise or outliers in training data (Alexander et al., 2011).

Although SVM is a relatively recent development in the context of remote sensing, it has been used successfully in many land cover classification studies (Huang, 2002; Foody and Mathur, 2004; Melgani and Bruzzone, 2004; Pal and Mather, 2005; Pal and Mather, 2006; Waske and Benediktsson, 2007; Rabe et al., 2010; Luo et al., 2010; Bilgin et al., 2011; Li et al., 2011; Longepe et al., 2011; Zhang et al., 2012; Li et al., 2012). In previous studies, SVM based classifiers performed more accurately than other classifiers or at least equally well. Because of its promising performance, researchers have been interested to incorporate SVM into existing classification schemes. In Moustakidis et al. (2012), an SVM based fuzzy decision tree was built for land cover classification of high spatial resolution hyper/multi-spectral images over urban and natural forest areas. In other studies, object-oriented and SVM classification were combined to separate different class information from multi-source image data or vast 3D LiDAR point clouds (Li et al., 2007; Zhan and Yu, 2011). The results showed SVM based classification providing high testing accuracies and low computational and storage demands, because SVM only considers support vectors close to the class boundary instead of whole training set.

For multi-class land cover classification over urban scenes, SVM approach has been applied on a variety of remote sensed data, such as hyperspectral imagery (Camps-Valls and Bruzzone, 2005; Zhang et al., 2012), full-waveform LiDAR (Mallet et al., 2008), and airborne LiDAR data (Secord and Zakhor, 2007; Samadzadegan et al., 2010a). Mallet et al. (2011) used SVM, with a feature selection step, to classify a full-waveform LiDAR data of an urban scene into three categories: building, vegetation, and ground. SVM achieved high testing accuracies in this study, underlining the high efficiency of SVM to discriminate complex urban landscape.

Lodha et al. (2006) classified airborne LiDAR data into four categories: building, tree, road, and grass using a SVM based classifier. With features of heigh, height variation, normal variation, and RGB, the classification results were proved stable and robust. In another study, a framework of urban change detection using satellite imagery and LiDAR data were presented (Malpica and Alonso, 2010), based on the SVM classification results. All the remote sensing data was resampled to 1 m resolution, and then the SVM based classifier successfully separated high vegetation and buildings. Features derived from grey level co-occurrence matrix (GLCM) method were introduced in Samadzadegan et al. (2010b) for LiDAR data classification using SVM. The urban areas were classified into building, tree, and ground with high accuracy. In other experiments, SVM was used for separating urban ground from other objects using LiDAR data and aerial imagery (Sarma and Yuan, 2009; Salah and Trinder, 2010).

Although SVM classifiers are effective and accurate in several studies, their parameters are still time-consuming and not easy to define (Waske et al., 2009c). Therefore, RF, a simple and fast classifier, was proposed and investigated in present study.

Multiple classifier system (MCS), also known as classifier ensemble, is to combine different classifier methods or variants of the same classifier approach. One of the latter type of MSCs, RF, a tree-based ensemble classifier introduced by Breiman (2001), has more recently been applied in a wide range of contexts. There have been several studies that focus on RF in the context of pattern recognition from remote sensing data, such as hyperspectral data (Ham et al., 2005), multisource remote sensing and geographic data (Gislason et al., 2006), SAR and optical

image (Waske and van der Linden, 2008), and multidate polarimetric synthetic aperture radar (Loosvelt et al., 2012).

According to Pal (2005), RF had excellent performance in the classification of multispectral data, comparable to SVM in terms of accuracy and training time, but it requires fewer parameters and the parameters are easier to define by user. Rodr ǵuez-Galiano et al. (2012) used RF classifier to map the complex Mediterranean land cover types by incorporating multitemporal Landsat TM imagery and digital terrain model variables. RF based classifiers showed superior performance compared with common classification methods, such as a simple decision tree, in a high dimensional feature space. When reducing the training sample size and increasing noise, RF is still robust in terms of overall accuracy. In another study, two different satellite sensors, IKONOS and QuickBird, were combined to test different pansharpening methods for urban area classification (Sveinsson et al., 2012). SVM and RF were used for classification. The results showed the RF classifier produced higher accuracies than the SVM method in almost every experiment.

Another approach to fused very high resolution (VHR) remote sensing imagery and LiDAR-derived digital surface models (DSM) was proposed for object-oriented landslides mapping using RF (Stumpf and Kerle, 2011). Yu et al. (2011) used the RF method to predict individual tree attributes based on both physical and statistical features derived from airborne LiDAR point cloud. The author concluded that RF technique is stable and reliable according to the experiment results. To classify urban scenes into four classes: building, vegetation, natural ground, and artificial ground, Chehata et al. (2009) applied RF with different features extracted

from full-waveform LiDAR and image data. It showed that RF classifier performed better than SVM by achieving higher classification accuracy and improving training and test computing time. However, very little research was focusing on urban land cover mapping by RF methods using airborne discrete LiDAR data and aerial imagery.

2.1.4 Knowledge-based Classification Technique

The knowledge-based classification (KBC) approach has increasingly been attractive in the field of land cover classification because of its ability of accommodating multi-source data (Kontoes and Rokos, 1996; Hung and Ridd, 2002; Thomas et al., 2003; Schmidt et al., 2004; Wentz et al., 2008; Ran et al., 2012). A KBC can consist of unsupervised and/or supervised classification methods. As a rule-based classification, generation of the rules is crucial. According to Hodgson et al. (2003), three common methods were proposed for building the rules in KBC methods. The first approach is explicitly eliciting knowledge and rules from experts and then refining the rules; the second one is implicitly extracting variables and rules using cognitive methods; and the last one is empirically generating rules from observed data and automatic induction methods.

However, the processing time of KBC rule generation is not satisfactory in certain applications. For example, the total classification time requirement by integration of LiDAR data and imagery is 9 hours when using KBC methods in Hodgson et al. (2003). It is much longer than 1.7 hours using unsupervised ISODATA classification and 0.3 hours using supervised maximum likelihood classifier (MLC). Using aerial imagery with LiDAR data, Huang et al.

(2008) developed a knowledge-based classification system (KBCS) to classify urban features into four categories: building, tree, grass, and road. Using a three-level-height classification rule-based method, different objects were extracted from low-, mid-, and high-height levels based on 15 rules associated with 18 numerical thresholds and 7 logical thresholds. The authors stated that the KBCS method receive higher accuracy than MLC and object-based classification. However, a common criticism of knowledge-based classification is that many thresholds involved and need to be assigned by human. In the experiment of Huang et al. (2008), only 2 thresholds were calculated by automatic optimum threshold selection method, the rest 23 thresholds (including 16 numerical and 7 logical thresholds) were input by user. The process of selecting thresholds is time-consuming and complicated. Also, the user is required to have sufficient experience and knowledge about urban-features extraction from remotely sensed data. Similar limitations were also found in Germaine and Hung (2011) where KBS methods were used to delineate impervious surface from multispectral imagery and LiDAR data.

2.2 Feature Selection Methods

In general, feature selection is defined as the task to remove irrelevant and/or redundant features (Guyon et al., 2006). Analysis of remote sensing data has recently shown more challenges and difficulties due to its larger volumes with a higher resolution and more feature bands. To reduce data dimensionality and computational complexity, unsupervised and supervised feature selection has drawn widespread attention in the field of remote sensing (Mitra et al., 2002; Liu and Yu, 2005). One of the most popular unsupervised techniques for feature reduction is principal components analysis (PCA). According to Lu et al. (2007), the massive majority of the image information can be retained while reducing the dimensionality through the accumulative eigenvalues of around 99% from PCA.

Several studies have proposed a variety of feature selection methods for SVM based classification (Frohlich et al., 2004; Chen and Lin, 2006; Chaelle and Keerthi, 2008). An evolutionary method, particle swarm optimization, was used for hyperspectral band selection to classify urban land cover (Yang et al., 2012). The results showed that this supervised feature selection method can greatly improve the SVM based classification accuracy, compared with other PCA methods. However, this feature selection approach requires the prior knowledge of the optimal number of features, which is not available in most cases. Huang and Wang (2006) proposed a generic algorithm (GA) based feature selection and parameters optimization approach for SVM classification. Nevertheless, the stochastic feature selection produced by GA may be

ineffective and time consuming facing high-dimensional remote sensing datasets (Stavrakoudis et al., 2012).

The nature of RF allows providing feature importance rankings as the criteria of feature selection, which is useful for multi-/hyperspectral data classification where the band dimensionality is high. It is of importance to investigate the influence of each predictive feature in the classification model for selecting the best features (Gislason et al., 2004; Gislason et al., 2006; Ghimire et al., 2010). Wang et al. (2009) applied a RF based feature selection for improving urban hyperspectral data classification. Martinuzzi et al. (2009) used RF based feature selection algorithm (varSelRF) and RF classifier to map snags and understory shrub species distribution from LiDAR data and inventory plots. The results showed agreement between the expectations from previous literature and the selected features with highest importance scores derived from RF. In another study (Ismail, 2009), HyMap data was utilised for detection and mapping of infested pine forests by varSelRF feature selection and RF classification. Overall the results showed the selected feature agreed the existing physiological knowledge.

There are two ways to evaluate feature selection methods (Liu and Motoda, 2008). One is to compare the results before and after feature section, including the number of selected features. The other is to compare two feature selection algorithms for the same task. Lu et al. (2007) applied two feature reduction methods, principal component analysis (PCA) and decision boundary feature extraction (DBFE) (Lee and Landgrebe, 1993), to a hyperspectral image covered a complicated agriculture area. The results of PCA or DBFE were inputted to two classifiers: MLC or extraction and classification of homogenous objects (ECHO) (Ketting and

Landgrebe, 1976). After comparing the classification results of PCA-MLC and PCA-ECHO with DBFE-MLC and DBFE-ECHO, the authors concluded that DBFE was better than PCA as the feature reduction method, yielding higher classification accuracy. Li et al. (2008) evaluated three typical feature selection methods by the same task of land cover classification using MLC and SVM classifiers. The classification accuracy along with the number of selected features was used as evaluation criteria. In another study, Laliberte et al. (2012) conducted a comparison of three feature selection methods for object-based classification, Jeffreys-Matusita distance (JM) (Swain and Davis, 1978), classification tree analysis (CTA) (Steinberg and Colla, 1997), and feature space optimization (FSO) (Definiens, 2009). The assessment criteria were ease of use, ability to rank and/or reduce input features, and classification accuracies. The results showed that JM required multiple steps, CTA had potential for overfitting decision tree, and FSO contained unclear feature ranking.

2.3 Chapter Summary

This chapter has presented an overview of the issues and methods of feature selection and classification for land cover mapping, especially for urban land cover classification using LiDAR data. To classify urban land cover types, a variety of classification methods has been reviewed. However, several major limitations have been found based on literature review: (1) some methods required prior knowledge on the study areas, (2) optimal parameters were difficult to find, (3) the processing time was not satisfactory, (4) many thresholds were involved and need to be assigned by human which was not user-friendly, (5) some methods entailed complex computational structure and required a large number of reliable training data, (6) classification methods such as DT had the over-fitting problem and required tree pruning after classification, (7) some of them showed poor accuracy performance. RF showed capability of solving all above obstacles, also embedded ability of feature selection. Up to now, there have been limited efforts to apply RF based classification method to LiDAR data fused with aerial imagery in urban areas. In the following chapters, feature selection and classification based on RF method is proposed for urban land cover classification from LiDAR data and aerial imagery. The results are also compared with other widely-used feature selection and classification methods.

Chapter 3
Introduction to Support Vector Machines and Random Forests Classifiers

In this chapter, a general guide to the concepts of SVM based and RF based classifiers are given. Section 3.1 provides the theoretical background and the development of SVM through different aspects. Section 3.2 describes the basic theories of building a random forest, and the parameter selection criteria. In the end, Section 3.3 summaries this chapter.

3.1 Support Vector Machines Classifier

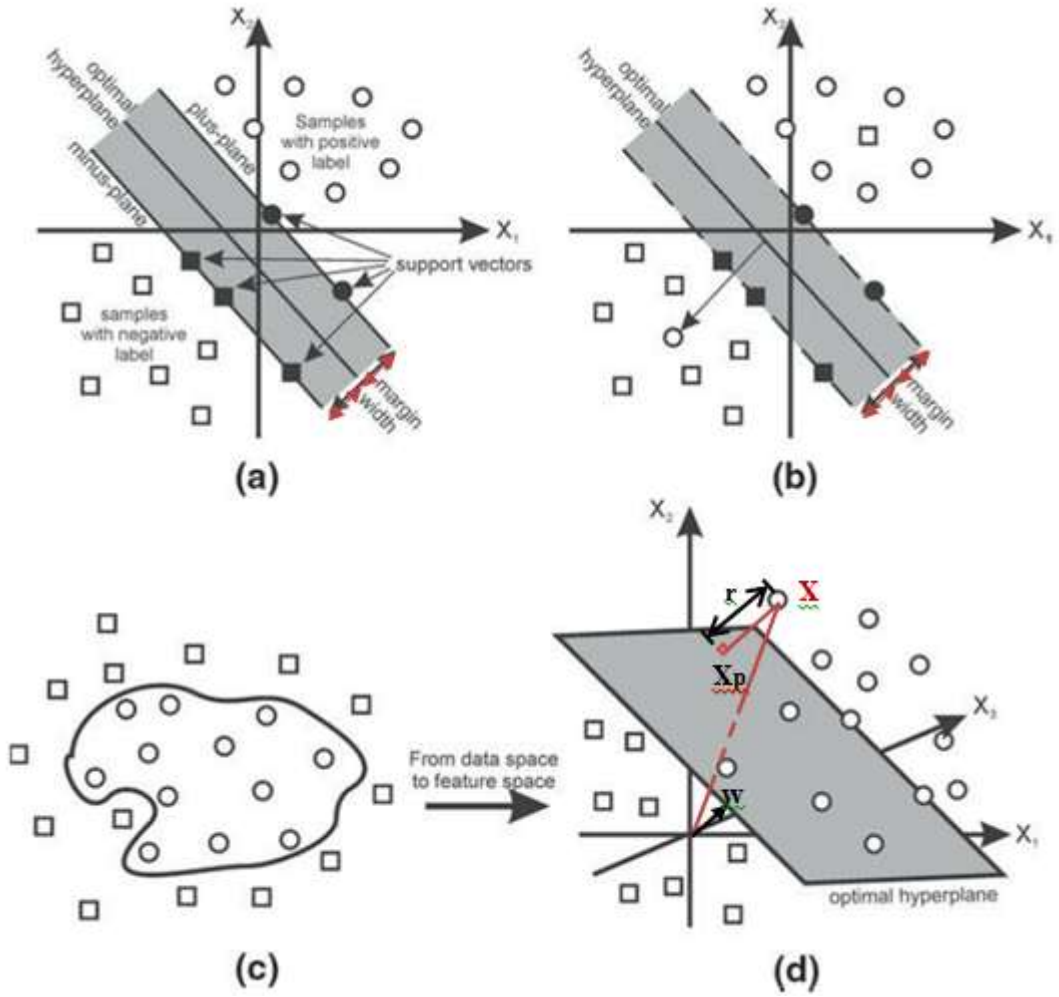


Figure 3.1 Support Vector Machine examples: (a) a linear binary SVM classifier, (b) a non-separable classification example using soft-margin, (c) a complex original data space is projected into a simple feature space shown as (d) using a kernel.

(Adapted: Yang, 2011a)

Considering only the training samples which are close to the class boundary, SVM has performed well in several multi-source data (including LiDAR data) classification problems,

even with small training sets (Dalponte et al., 2008; Jones et al., 2010; Trinder and Salah, 2011). This section will first explain the principles at the core of the SVM classifier. Afterwards the parameter selection strategies and application software are also introduced.

The original formulation of SVM was introduced in Vapnik (1979), and aims to construct an optimal separation hyperplane that discriminate the dataset into discrete predefined number of classes by using a training sample subset as the support vectors. In the learning step, a classifier with maximum margin which defined by support vectors is iteratively searched within a multi-dimensional feature space to classify the training samples, and then to classify the test samples. Margin refers to the distance from the hyperplane to the samples closest to it on either side. The optimal separation hyperplane is the one that obtains the largest margin. An important generalization aspect of SVM is that the number of support vectors is frequently smaller than the size of available training samples. Unlike other classifiers, which directly deliver a class label as final output (e.g., decision tree) and calculate probabilities of class memberships (e.g., MLC); SVM provides distances of each input vector to the optimal hyperplane (Waske et al., 2009a).

Figure 3.1(a) illustrates the simplest form of SVM, a linear binary classifier, which assign a given test sample a label from two class memberships: positive and negative. The training sample can be mathematically presented as $X = \{(x_i, d_i)\}_{i=1}^{N}$; where x_i is an input sample, N is the size of the training samples, and d_i is a class label: $d_i = +1$ if $x_i \in C_1$ and $d_i = -1$

if $x_i \in C_2$. The problem is eventually converted to find the parallel hyperplane described as (Burges, 1998):

$$w^T x_i + b \geq +1 \text{ for } d_i = +1 \text{ (plus-plane)} \tag{3.1}$$

$$w^T x_i + b \leq -1 \text{ for } d_i = -1 \text{ (minus-plane)} \tag{3.2}$$

where w is the weight vector and b is the bias parameter.

The Equations (3.1) and (3.2) can be combined and rewritten as

$$d_i(w^T x_i + b) \geq +1 \tag{3.3}$$

As illustrated in Figure 3.1(d), x_{pi} is the normal projection of x_i onto the hyperplane H, r is the distance vector of x_i to H, and w is the normal vector of H. Therefore, x_i can be expressed as (Duda et al., 2001)

$$x_i = x_{pi} + r\frac{w}{\|w\|} \tag{3.4}$$

Because $w^T x_{pi} + b = 0$, then

$$r = \frac{w^T x_i + b}{\|w\|} \tag{3.5}$$

The Equation (3.5) implies that $\frac{b}{\|w\|}$ is equal to the distance vector from the origin to H, so b is the key factor for the location of hyperplane. In order to find the maximum margin, the absolute value of r should be at least some value p, then

$$\frac{d_i(w^T x_i + b)}{\|w\|} \geq p, \forall i \tag{3.6}$$

To find a unique solution for maximum$p$, the constraint $p\|w\| = 1$ is imposed, and then the goal is defined as (Cortes and Vapnik, 1995)

$$\min \frac{1}{2}\|w\|^2 \text{ subject to } d_i(w^T x_i + b) \geq +1, \forall i \tag{3.7}$$

In solving this standard quadratic optimization problem, Equation (3.7) is converted to an unconstrained problem by means of Lagrange multipliers $\propto_i$ as maximizing (Cristianini and Shawe-Taylor, 2000)

$$\mathrm{L}(\propto) = \sum_{i=1}^{N} \propto_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \propto_i \propto_j d_i d_j x_i^T x_j \tag{3.8}$$

Subject to $\sum_{i=1}^{N} \propto_i d_i = 0$, and $\propto_i \geq 0, \forall i$.

After solving for $\propto_i$, the support vectors are obtained by selecting the set of x_i whose $\propto_i > 0$. Then w and b can be calculated based on (Alpaydin, 2004)

$$w = \sum_{i=1}^{N} \propto_i d_i x_i \text{ , and } b = d_i - w^T x_i \quad (3.9)$$

respectively.

So far, all the problems have been treated as linear classification task. To deal with the nonseparable cases as shown in Figure 3.1(b), Cortes and Vapnik (1995) introduced a soft margin method with a set of slack variables, $\xi_i \geq 0$, which measures the deviation from the margin. Then Equation (3.3) can be reformulated as

$$d_i(w^T x_i + b) \geq 1 - \xi_i \quad (3.10)$$

There are three possible situations: if $\xi_i = 0$, x_i is correctly separated and classified; if $0 < \xi_i < 1$, x_i is correctly labeled but lie in the margin; if $\xi_i \geq 1$, x_i is misclassified. Similarly, Equation (3.8) is refined as

$$\mathrm{L}(\propto) = \sum_{i=1}^{N} \propto_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \propto_i \propto_j d_i d_j x_i^T x_j \quad (3.11)$$

Subject to $\sum_{i=1}^{N} \propto_i d_i = 0$, and $0 \leq \propto_i \leq \mathrm{C}, \forall i$,

where C is an error penalty factor which controls the trade-off between the number of support vectors and the number of the non-separable samples.

Boser et al. (1992) introduced a kernel trick to solve the non-linear classification problem. The complex problem in the data space is mapped to a new high-dimensional feature space

where a linear model would be created, as illustrated in Figures 3.1(c) and 3.1(d), using suitable nonlinear transformation functions.

Only the kernels meet Mercer's theorem (Mercer, 1909) can be used in SVM, the most common ones are linear, Gaussian radial-basis function (RBF), polynomial, and sigmoid kernels, and their mathematical representations are (Fletcher, 2009; Haykin, 1999)

$$\text{Linear: } K(x, x_i) = x^T x_i \tag{3.12}$$

$$\text{RBF: } K(x, x_i) = exp\left(-\frac{\|x - x_i\|^2}{\gamma^2}\right), \gamma > 0 \tag{3.13}$$

$$\text{Polynomia: } K(x, x_i) = (cx^T x_i + r)^\rho, \gamma > 0 \tag{3.14}$$

$$\text{Sigmoid: } K(x, x_i) = \tanh(\gamma x^T x_i + r) \tag{3.15}$$

where x is a sample in the original data space, x_i is an corresponding sample in the feature space, γ, ρ, and r are user-defined kernel parameters.

In general applications, including remote sensing data analysis, the RBF kernel is a widely used kernel function due to its higher performance and practical conveniences (Gómez-Chova et al., 2011). The linear kernel is only a special case of RBF (Keerthi and Lin, 2003), and the sigmoid kernel with certain parameters behaves like RBF (Lin and Lin, 2003). Furthermore, Compare to RBF, the polynomial kernel has higher model complexity and longer processing time

because it requires more hyperparameters. However, if the number of features is huge, the linear kernel may be the better choice (Hsu et al., 2010).

SVM based classifier was originally invented to solve binary classification problems as described above. However, remote sensing applications usually need to discriminate among a broad range of land-cover or land-use classes. Thus two approaches have been developed for SVM multiclass classification (Weston and Watkins, 1999):

- One-against-all method: it is probably the earliest used strategy for SVM multiclass classification problem, for example in Bottou et al. (1994). For an N-class problem, N binary classifications are constructed and trained to separate each class from all remaining classes.

- One-against-one method: it was introduced by Knerr et al. (1990), and then first used on SVM by Friedman (1996) and Kreßel (1999). Each pair of classes is calculated as a binary classification problem. Finally, all N(N-1)/2 binary classification results are combined, and a majority voting is conducted to generate the final output. Each sample in the final output is labeled with the class having most votes. Instead of simple voting method, Wu et al. (2004) introduced pairwise coupling to obtain SVM multi-class classification decisions and found that it is more stable. This method is suitable for solving the classification problem with large amount of data.

Samadzadegan et al. (2010b) concludes that one-against-all derived higher accuracy, while one-against-one is much less time consuming. For dealing with larger dataset, one-against-one method is a better choice than the other.

Because of its high performance and processing efficiency as mentioned above, RBF kernels are chose to serve in the implementation of SVM classification in this study. According to Equations (3.11) and (3.13), two parameters are needed to be specified by user: the penalty factor C and the width of the kernel function γ.

The best C and γ are not known beforehand for the current problem, thus a parameter selection has to be done. In this study, k-fold cross-validation is performed to find the pair of optimal parameters (C, γ) from an infinite number of possible ones. This procedure is beneficial to prevent the overfitting problem (Hsu et al., 2010). First, the training dataset is divided into k subsets of equal size. Then each subset is tested using the classifier trained on the remaining k-1 subsets. Consequently each subset in the training dataset is tested once, so the percentage of correctly classified samples is calculated at each time (k times in total) as the cross-validation accuracy (CVA) (Secord and Zakhor, 2007).

A grid search is recommended to executed in parallel to pick the (C, γ) with the highest CVA (Hsu et al., 2010). To reduce the computation time, the practical way is testing exponentially growing sequences of (C, γ) values using a coarse grid first, e.g., $C = 2^{-5} - 2^{15}$, $\gamma = 2^{-15} - 2^{3}$, multiplier as 2^{2} ,followed by a finer grid (multiplier as $2^{0.25}$) in the

neighborhood of best (C, γ) which derived from the previous coarse grid search (Hsu et al., 2010). Once (C, γ) have selected, they are used to train the whole training dataset to determine the final classifier with the optimal hyperplane.

3.2 Random Forests Classifier

As indicated in the above section, SVM approach is based on a relative complex decision boundary which is computed after transferring the original linear nonseparable samples into a higher dimensional feature space. In contrast, RF approach replies on different decision tree classifiers with many rather simple decision boundaries paralleling to the original feature axis. Feature space transformation is not needed in an RF classifier, so it could save much more computing time than a SVM classifier. Several studies have shown the advantages of RF in the field of land cover classification (Pal, 2005; Gislason et al., 2006; Smith, 2010; Zhu et al., 2011). In the rest of this section, three aspects of RF will be presented: the basic principles, the parameters determination and algorithm implementation tools.

The RF classifier developed by Breiman (2001) is a combination of decision trees $\{DT(x, \Theta_k)\}_{k=1}^{T}$, where x is an input vector, and Θ_k denotes a random vector which is sampled independently but with the same distribution as the past $\Theta_1, \ldots, \Theta_{k-1}$. T bootstrap samples are first drawn from the training data, and then an un-pruned classification and regression tree (CART) is grew from each bootstrap sample β where only one of *M* randomly selected features is chose for the split at each node of CART.

Finally, classification output is created based on a majority vote of the predictions from all individually trained trees. The workflow is shown as Figure 3.2.

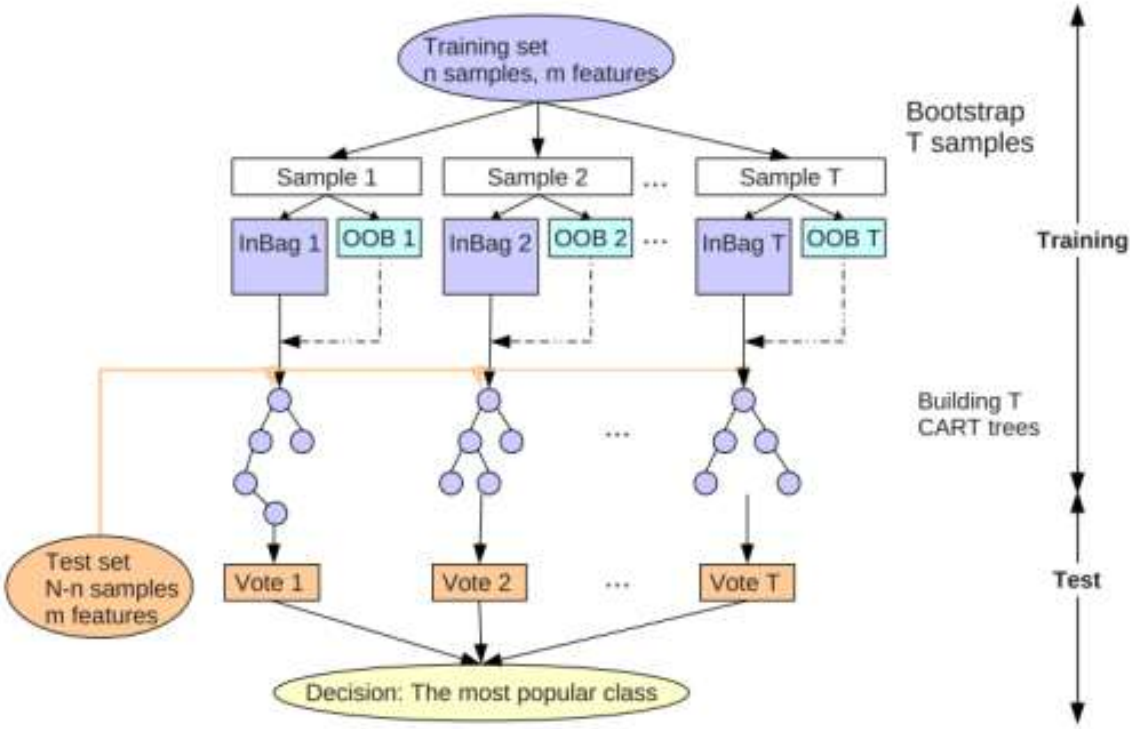


Figure 3.2 Random Forests classification workflow

(Source: Guo et al., 2011)

A bootstrap dataset is a set of randomly selected points which are drawn with replacement from the training dataset (Duda et al., 2001). To obtain the same size with the original training sample, a bootstrap dataset nearly always has duplication of individual points. While each tree in the forest is constructed using a different bootstrap dataset, about one-third of the points are left out of each bootstrap sample, called Out-of-Bag (OOB) data. On the average, each training point would be out-of-bag around 36% of the times. Thus, an estimate of the classification error rate can be derived based on the training data. The OOB samples are run down through the trees, and

count the times of incorrect classification which are then averaged by all the trees yielding the total OOB error rate. It is unbiased (Breiman, 2001) and can be used to plot the relationship between the OOB error and the number of trees. The number of trees should be enough while the error rate becomes stable (Horning, 2010).

There are two additional indications that are produced from the RF: proximities analysis and variable importance. The former one measures the distance from one sample to others, which can be used as outlier detector (Joelsson et al., 2008). It based on the idea that similar samples should fall in the same terminal nodes more frequently than dissimilar ones (Liaw and Wiener, 2002). The proximity measure also can be used to visualize the structure of high dimensional data (Breiman, 2003). Variable importance measures the importance of the predictor variables (features). To estimate a feature importance, the OOB samples are first run through the trees and count the votes for the correct classification. Then, the prediction accuracy is repeatedly obtained after randomly permuting all the values of this feature while all the other features stay the same. The importance score is the decrease of the correct class votes after the variable permutation, averaged over all the trees. The intuition is that a random variable permutation can simulates the absence of that variable from the forest (Guo et al., 2011). Thus the higher an average accuracy decrease is, the more important that feature is. Another index to measure importance is the Gini impurity which can be written as (Breiman et al., 1984)

$$\text{Gini}(\beta) = \sum \sum (f(C_i, \beta)/|\beta|)\,(f(C_j, \beta)/|\beta|) \tag{3.16}$$

where $f(C_i, \beta)/|\beta|$ is the probability that the randomly selected pixel belongs to class C_i.

By summing up the Gini impurity decreases for each variable in the random forest, it provides a fast way to measure variable importance which is usually consistent with the previous permutation method (Breiman, 2003).

In a RF based classification, two parameters need to be specified by user: the size *M* of a random subset of features and the number of trees *T*. The selection of parameter *M* has influence on the final error rate. If *M* is increased, both the correlation between the trees and the strength (classification accuracy) of individual tree in the forest are increased. The error rate is proportional to the correlation, but inverse proportional to the strength (Joelsson et al., 2008). Usually, *M* is set to the square root of number of features (Gislason et al., 2006). Because RF is fast and not overfit, the number of trees *T* can be as many as possible. However, due to the memory limit of the machine, *T* is usually several hundred to thousand (Horning, 2010).

3.3 Chapter Summary

This chapter presented the theoretical background of SVM and RF. Each technology was presented through two major phases: the basic principles and the parameters determination. SVM based classifier aims at the definition of an optimal separation hyperplane to separate different classes in a multi-dimensional feature space mapped by a kernel function such as RBF. The required parameter varies when adopting different kernel functions. The common methods used to find best parameters of SVM are grid search and cross validation. RF classifier, one of the classifier ensemble methods, offers an elegant fashion to build a strong classifier by combining a set of independent weak DTs. Usually, RF classifier only requires two parameters which are user-friendly defined.

Chapter 4
Methodology

In this chapter, the methodologies of this study are presented. Section 4.1 provides the information on the study area and datasets. Section 4.2 lists and explains all the input features derived from LiDAR and aerial imagery. Section 4.3 presents the methods to collect the training and reference data. Section 4.4 introduces the software packages and parameters used in this thesis for feature selection and urban land cover classification, followed by an introduction of evaluation methods in Section 4.5. Last, Section 4.6 presents the summary of this chapter.

4.1 Overview of the Proposed Methodology

A general overview of the methodology in this thesis is shown in Figure 4.1. First, proper features were extracted from LiDAR data and aerial imagery, organizing into three groups: LiDAR feature group, image feature group, and all feature group. Next, feature selection was applied to the previous generated feature groups based on PCA and RF techniques. Afterward, the supervised classification with different training sample size was performed on original features (derived from feature extraction step) and reduced features (derived from feature selection step), according to three classifiers: MLC, SVM, and RF. Finally, the study area was classified into four major urban land cover categories. The classification results were compared with reference data to evaluate the classification schemes.

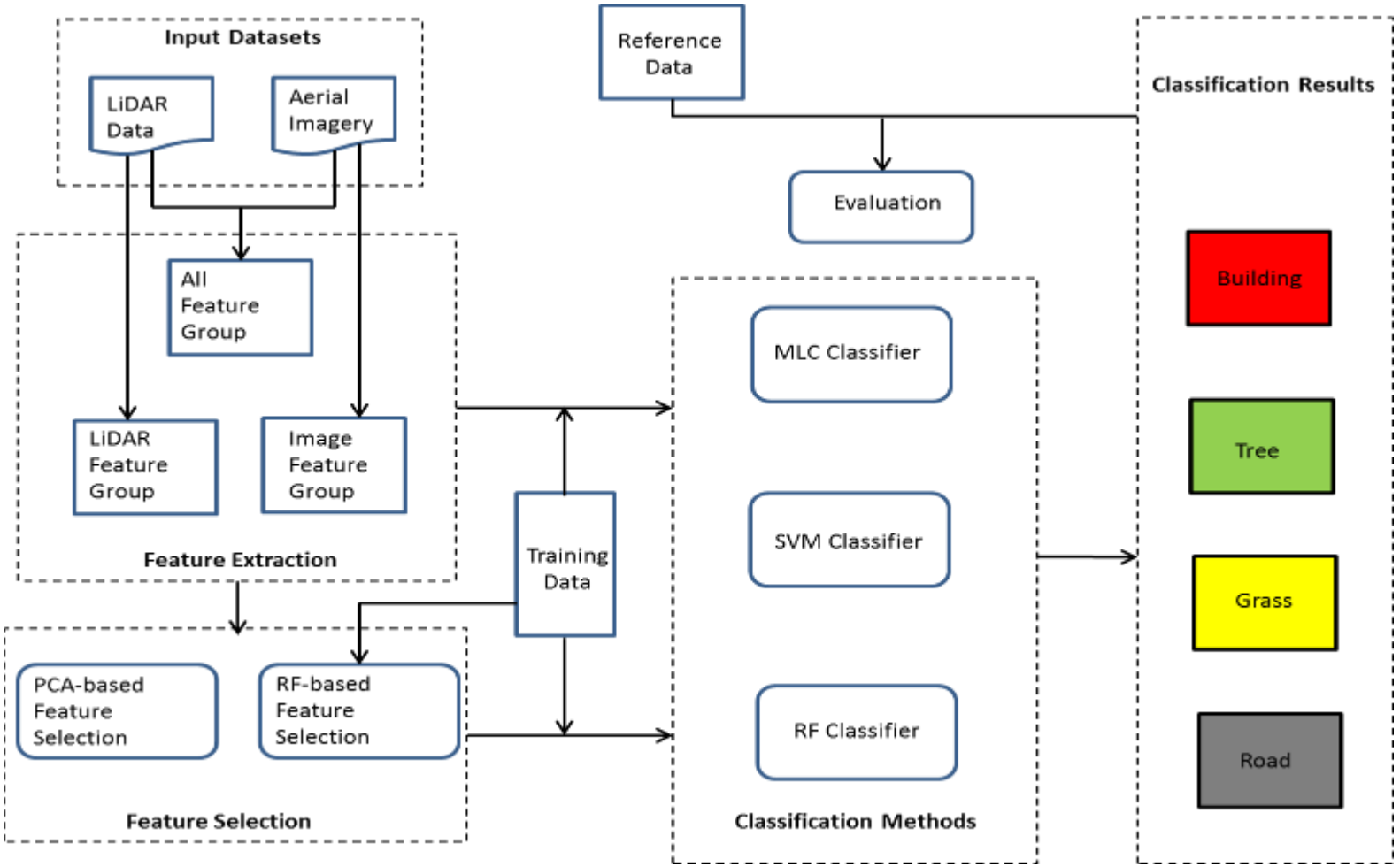


Figure 4.1 Overall workflow of this study

4.2 Study Area and Datasets

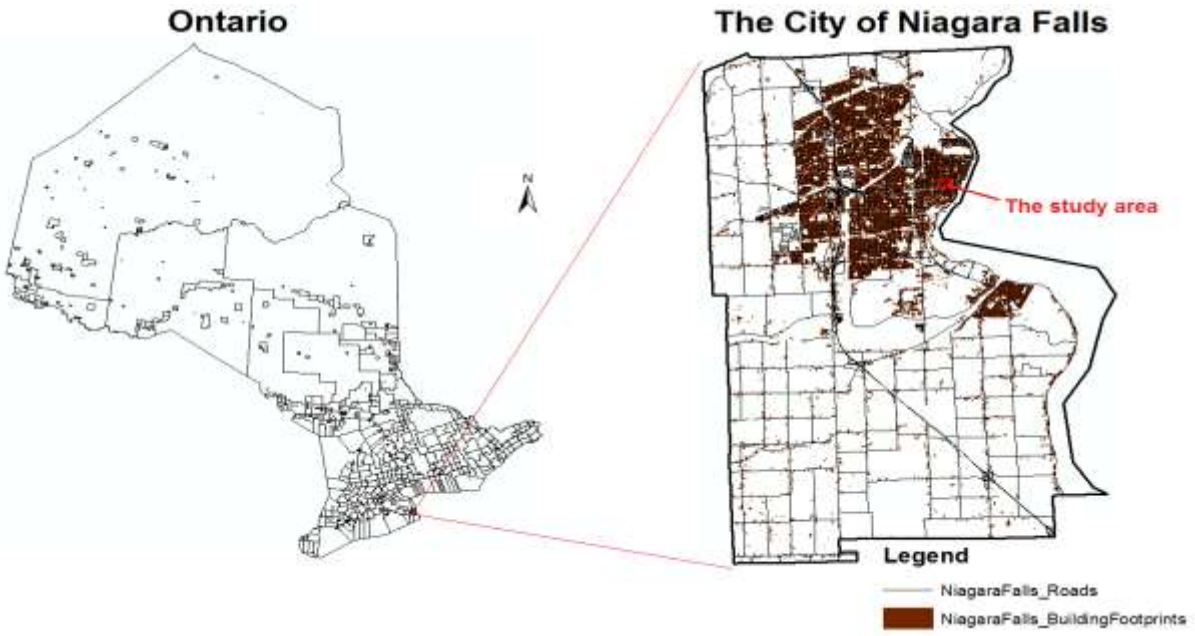
The City of Niagara Falls, in the Golden Horseshoe region of southern Ontario, is part of the Regional Municipality of Niagara, Canada's 12th largest Census Metropolitan Area (CMA). It has a population of 82,184 inhabitants and population density of 392.1 per km ^{2} (Statistics Canada, 2007). In this thesis, an urban area of the City of Niagara Falls was selected as a test site (see Figure 4.2). The almost flat region contains a college school, a shopping plaza, and more than 300 residential and business buildings, according to the GIS building footprint data. Land cover components are typical of those in urban and suburban scenes, including houses with both flat and pitched rooftops, impervious concrete and asphalt surfaces such as parking lots, sidewalks, and roadways, and pervious vegetation surfaces such as trees and grass. The great variety of the land cover in the study area makes it ideal for this study. The desired map classes for this project are detailed in Table 4-1. Road includes road, parking lot, and street items such as car and traffic light. Grass includes green and dry grass. Tree includes coniferous and deciduous trees.

Table 4-1 Land cover types in the study area

Pervious Classes	Impervious Classes
Tree	Building
Grass	Road

The LiDAR data were acquired in 2004 by an Optech ALTM 3100 system. The airplane flew at an average height of 1,190 m above the mean sea level, with a DSS 301 SN0039 camera on board for the 0.5m-resolution aerial photographs with RGB bands. The horizontal and vertical accuracies are 0.6m and 0.15m (1 sigma standard deviation), respectively. The average point density and point spacing within the test site is about 4 points/m^2 and 0.5 m, respectively. The LiDAR dataset records both range (first- and last-returns) and intensity information of the laser pulse which are co-registered with the 0.5m-resolution aerial photographs containing 803,439 pixels.

A grey-scale orthotimage covering the study area was taken in the spring of 2006, with a 10cm resolution, is available to use in the reference data generation step. Several GIS vector dataset including building footprints and roads are also available. All the experiments were conducted on a consumer level laptop (Intel Core i7 2.67GHz CPU, 4G RAM, 64-bit Operating System).



(a) Location of the study area

(b) Aerial image of the study area

Figure 4.2. Study area

4.3 Feature Groups

A set of feature vectors which is focused on discriminating between classes is crucial to land cover classification (Wang et al., 2009). For urban land-cover classification in this study, a total of 31 features were generated from LiDAR height information, reflection intensity sensor and aerial imagery, which are listed in Table 4-2. The features were grouped into image feature group, LiDAR feature group, and all feature group.

Table 4-2 Feature groups used in this study

Feature group	Feature extraction methods	Features
Image feature group	RGB bands	Red, Green, Blue
	GLCM based features	Red_Contrast, Red_Entropy, Red_Correlation, Green_Contrast, Green_Entropy, Green_Correlation, Blue_Contrast, Blue_Entropy, Blue_Correlation
LiDAR feature group	Height based features	DSM, fr-lr, HVar, HDiff
	LiDAR intensity	Intensity
	GLCM based features	DSM_Contrast, DSM_Entropy, DSM_Correlation, Intensity_Contrast, Intensity_Entropy, Intensity_Correlation
	Eigenvalue based features	Eigen1, Eigen2, Eigen3, Anisotropy, Planarity, Linearity, Sphericity
All feature group	LiDAR-TVI & LiDAR feature group & image feature group	LiDAR-TVI, DSM, fr-lr, HVar, HDiff, Intensity, DSM_Contrast, DSM_Entropy, DSM_Correlation, Intensity_Contrast, Intensity_Entropy, Intensity_Correlation, Eigen1, Eigen2, Eigen3, Anisotropy, Planarity, Linearity, Sphericity, Red, Green, Blue, Red_Contrast, Red_Entropy, Red_Correlation, Green_Contrast, Green_Entropy, Green_Correlation, Blue_Contrast, Blue_Entropy, Blue_Correlation

- Height-based features

The elevation image of first return pulses is generated to serve as a height value feature during the classification. It distinguishes elevated objects such as buildings and trees from others. In addition to DSM image, the height differences between first return and last return pulses of LiDAR points are rasterized as a feature layer. It has the capability to detect trees due to the penetrability of canopy (Rottensteiner and Clode, 2009).

Height difference (HDiff) and height variance (HVar) were calculated to describe the local height variation. HDiff is the difference between maximum and minimum height value in a 3*3 neighborhood, while HVar is the variance of the height values in the same neighboring region. These two features have the potential to separate "rough" surface (e.g., trees) with other "smooth" objects, such as rooftops and grass (Huang et al., 2011).

- LiDAR reflection intensity

Intensity, for most LiDAR systems, refers to the ratio of return pulse energy to transmitted pulse energy. It is affected by variable factors, such as roughness and orientation of object surface, variations in path length, laser beam divergence, structure and density of object, background saturation, and signal attenuation through the atmosphere (Starek et al., 2007). LiDAR intensity values have been calibrated and adjusted before delivery to reduce the variations and systematic errors derived by spherical loss, topographic and atmospheric effects. After calibration and correction,

intensity values show similarity in the homogeneous surface. Usually, standard deviation, median, and mean values of the homogeneous surface are used for evaluation of correction, along with the visual inspection (Hofle and Pfeifer, 2007; Oh, 2010). Intensity can be used to improve the land cover classification when it is difficult to distinguish different objects using only elevation information (Li et al., 2008).

- RGB bands

From aerial image, three separated feature layers, Red, Green, and Blue, are directly extracted. They represent the responses of all objects on the terrain surface to visible light. Usually, they preserve the corner and boundary information better than LiDAR data (Zhang, 2010).

- GLCM-based features

Textural features are derived from GLCM (Haralick et al., 1973). According to Hall-Beyer (2008), GLCM texture features can be classified into three groups: contrast group, orderliness group, and descriptive statistics group. Selection of at least one feature from each group is suggested by Ozdemir and Karnieli (2011), because of the strongly correlation inside each group. Only three features (Contrast, Entropy, and Correlation) are selected to use in this research (Clausi, 2002). Contrast from contrast group refers to the amount of local variations; Entropy from orderliness group is a measure of the

amount of disorder; Correlation from descriptive statistics group measures the linear dependencies of grey levels (Haralick et al., 1973). 5 bands were used to generate GLCM-based texture features: Red, Green, Blue, Intensity, and DSM. All the GLCM-based features were generated with a window of 3×3 pixels by the build-in 'Co-occurrence Measues' function in ENVI. Each time, the texture values were put in the central pixel of the 3×3 window. At last, the values for the edge pixels of whole image were interpolated from the nearest texture values.

- Eigenvalue-based LiDAR features

This group of features describes the spatial local distribution of LiDAR points, which are able to detect corners, planes, volumes, and lines (Gross and Thoennessen, 2006):

$$Anisotropy = \frac{\lambda_1 - \lambda_3}{\lambda_1} \tag{4.1}$$

$$Planarity = \frac{\lambda_2 - \lambda_3}{\lambda_1} \tag{4.2}$$

$$Sphericity = \frac{\lambda_3}{\lambda_1} \tag{4.3}$$

$$Linearity = \frac{\lambda_1 - \lambda_2}{\lambda_1} \tag{4.4}$$

where $\lambda_1 > \lambda_2 > \lambda_3$ are the eigenvalues computed from the variance-covariance matrix within the local region. These three eigenvalues are also used as features (Chehata et al., 2009).

- Image and LiDAR derived feature

To discriminate vegetation types, NDVI (normalized difference vegetation index) calculated from the red and near-infrared (NIR) bands was widely adopted and applied to the remote sensing images. Deering et al. (1975) proposed the TVI (transformed vegetation index) to modify the NDVI from the Poisson distribution to the normal distribution, by adding 0.5 to the NDVI values and taking the square root of the results. However, the NIR band is not always available. Recently, LiDAR- TVI is introduced in Huang et al. (2008) for discrimination of grass and tree, especially performs well for grass. It is the square root of pseudo-NDVI which is calculated by replacing the NIR band with LiDAR Intensity (wavelength: 1064 nm) (Rottensteiner et al., 2005; García-Gutiérrez et al., 2010). A constant of 0.5 was added to eliminate a negative value, and the constant varies with minimum pseudo-NDVI value:

$$\text{lidar} - \text{TVI} = \sqrt{\frac{(\text{lidar intensity} - \text{red})}{(\text{lidar intensity} + \text{red})} + 0.5} \qquad (4.5)$$

Sufficient feature layers were generated and then used in the following classification procedures.

4.4 Training and Reference Data

Both training and reference data are indispensable in the supervised classification schemes. Training samples, as input to yield a classification map, represent the spectral signatures of each class; and reference data is required to systematically compare with the classification results for further accuracy assessment. Training and reference data for the present study were collected from the study area in Section 3.1. Different selected sampling sizes and schemes are introduced in this section.

There are many ways to collect ground truth information for supervised classification. For example: in situ information collection, on-screen selection of regions of interest (ROIs), and/or on-screen seeding of training data. In an ideal situation, the centroid coordinates of all the samples are measured using a global positioning system (GPS) receiver by visiting them in the field, and observations of ground information are very carefully conducted at the same time. Unfortunately, sometimes part of the locations may be inaccessible because of private land owners, extremely rugged terrain, or other regulations. And due to the lack of high-accuracy GPS equipment, i.e. real time kinematic (RTK) GPS receivers, this field-work method is not adopted in this thesis. To collect the ground truth samples for this research, 6500 pixels were randomly selected by "Create Random Points" tool in ArcGIS 10 with a minimum allowed distance of 1 m. Then, each location was investigated and labeled using higher-spatial-resolution remotely sensed data (10 cm resolution orthoimage) based on human visual interpretation. To reduce the effects of geometric misregistration, 8 pixels surrounding the reference pixel

were investigated, and then the class with the highest frequency of occurrence was assigned to the reference pixel (Jensen, 2005).

An equalized random sampling was performed, guaranteeing that all classes are equally distributed in the training set. The required size of the training dataset is commonly estimated from either basic sampling theory or simple heuristics method (Foody and Mathur, 2006). The former is based on the assumption that the spectral response of a class follows a normal distribution, and the total number of pixels in a remote sensing image dataset is generally large. Thus, the training sample size, n, could be determined as (Foody et al., 2006)

$$n = \frac{\sigma^2 z^2}{h^2} \tag{4.6}$$

where σ is the planning or estimated value for the population standard deviation of the distribution, z is the value of the z score at a specified level of confidence, h is the specified half-width of the confidence interval.

Alternatively, simple heuristic method is often used to specify the training sample size as at least 10-30N_f per class, where N_f is the number of spectral wavebands or other discriminating features used in the classification (Foody, 2009). In this research, four class memberships are derived using multiple feature layers which will be explained in next section. To investigate the possible effect of the number of training samples on the classification accuracy, training data with different size were created, containing 250,

500, and 1000 training samples per class, respectively (from now on referred to as tr#250, tr#500, and tr#1000).

All the training data were excluded to perform the error evaluation, in this way, biased accuracy assessment was avoided. To populate an error matrix and preform an accuracy assessment in this research, an independent reference dataset was available containing 2500 samples. Since multiple classes are involved in the classification, the reference sample size (N) is determined based on multinomial distribution equation (Tortora, 1978):

$$\mathrm{N} = \frac{B\Pi_i(1-\Pi_i)}{b_i^2} \tag{4.7}$$

where $\Pi_i$is the percentage of the sample size in the ith class which occupies closest to 50% of all the k classes over study area, b_i is the anticipated precision, B is determined from the χ^2(Chi-square) distribution table with 1 degree of freedom and $1-{}^{\alpha}/_{k}$, $(1-\alpha)$ is confidence interval.

However, in present research the true proportion of any of the 4 classes is unknown. So the worst-case multinomial distribution algorithm was adopted to assume that one class occupies 50% of the test site, *b*=5%, confidence interval is set to 95%. According to Equation (4.7), at least 624 reference samples should be collected to populate an error matrix. In present study, the sample size was future extended to 2500.

4.5 Feature Selection and Classification

PCA based feature reduction was applied in ENVI to retain around 99% accumulative eigenvalues for each multi-band image with different feature groups. At the RF based feature selection stage, the R package varSelRF was utilized to iteratively eliminate 20% of the least important features. The selected set of features is the one with the minimum OOB error rate. For each RF based feature selection, the package varSelRF was iterated 10 times, and then the most frequent selected feature combinations were used in the following classification step. In total, 81 classification experiments were conducted in this study. Three feature groups (including image feature group, LiDAR feature group, and all feature group) were used as input to nine classification schemes (including MLC, SVM, RF, PCA-MLC, PCA-SVM, PCA-RF, RF-MLC, RF-SVM, and RF-RF), along with three training sample sizes (including tr#250, tr#500, and tr#1000).

The one-against-one SVM classifications were implemented using imageSVM (Janz et al., 2007), which is an open-source IDL/ENVI plug-in for SVM classification and regression analysis of remotely sensed data with common formats. It is developed by the Geomatics Lab of Humboldt-Universität zu Berlin based on the widely accepted LIBSVM approach (Chang and Lin, 2011). For SVM classification, it comprises three steps: image scaling, automatic or user-defined parameterization based on training data, and classification of input image data. In the automatic parameterization setting, the RBF kernel parameter γ and regularization parameter C are selected using grid search and 3-

fold cross validation. First, a coarse grid search was applied within ($C = 2^0 - 2^{10}$, $\gamma = 2^{-7} - 2^3$) with a multiplier of 2^2, and then a finer grid with multiplier of $2^{0.25}$ was built to find the final parameters in the neighborhood of best (C, γ) which derived from the previous coarse grid search. Beside the final classification image, imageSVM also offers the probability image of each class as output based on Platt (2000) and Wu et al. (2004). The imageSVM program is more user-friendly than the build-in SVM program in ENVI, due to the availability of the function of automatic parameterization.

In addition, the R 'randomForest' package was used to run RF classifications. The randomForest package for the R statistics language (R Development Core Team, 2009) is a common open source tool for full-featured implementation including feature importance calculation. It is an R interface to the Fortran programs by Breiman and Cutler (available at http://www.stat.berkeley.edu/users/breiman/). R supports a variety of geospatial data analysis packages for handling of GIS data and remote-sensing images, such as *maptools* (Lewin-Koh and Bivand, 2011) and *sp* (Pebesma and Bivand 2005; Bivand et al., 2008). For selecting the size *M* of a random subset of features for RF, the square root of number of features, half of it, and twice of it were used to generate a random forest, and the one produced the lower OOB error rate was picked. In this study, the OOB error rate did not change dramatically when using different *M*, and the square root of number of features was used. For selecting number of trees, different values were

tested, and the one showed effective and stable results was used. In this study, it showed that 1000 trees were enough for all the RF classifications.

4.6 Evaluation Methods

No classification is complete until a valid evaluation on its result map has been done (Foody, 2004b). On the one hand, map users need to know if the map quality is suitable for a particular purpose, and on the other hand map producers need to evaluate the mapping process and further improve it. To determine the accuracy of classification maps derived from the previous classifiers, error matrix (Congalton, 1981), was built for each map, and then overall accuracy, User's accuracy, Producer's accuracy were obtained followed by a Kappa analysis. After evaluation of individual classifier, several statistical based techniques were applied for comparison of the produced maps. Besides the statistical accuracy assessment methods, a visual interpretation was also applied for map evaluation.

Error matrix, also known as confusion matrix, was utilized in this research for accuracy assessment of the classification results. Each column in the matrix indicates the reference data, while each row represents the predicted class (Congalton and Green, 2009). The number of classification result in a certain category is related to the number of reference data in a particular category. All the diagonal numbers refer to the correctly classified pixels of different class. The major advantage of error matrix is the effective

representation of individual accuracies for each class along with commission errors and omission errors. A commission error occurred when a classified pixel is included in a category to which it does not belong. An omission error occurred when a classified pixel is excluded from a category to which it belongs. Beside plainly showing commission and omission errors, the error matrix can be used to generate descriptive evaluations such as overall accuracy, Producer's accuracy, and User's accuracy.

The overall accuracy is defined as the sum of the major diagonal divided by the total number of pixels in the error matrix (Congalton and Green, 2009). Producer's and User's accuracies introduced by Story and Congalton (1986) are two measures of omission and commission errors. For each row, the Producer's accuracy is calculated by dividing the number of correctly classified pixels in that row by the total number of pixels in the row. User's accuracy is similar, but it is computed as the percentage of the number of correctly classified pixels in each column to the total number of pixels in that column (Congalton and Green, 2009). Kappa analysis is another important technique in accuracy assessment, was first introduced to the remote sensing field by Congalton (1981). It derives a Khat (coefficient of agreement) which measures the accuracy between classification result and reference data using the major diagonal and the chance agreement. It can be represented as (Jensen, 2005)

$$\hat{K} = \frac{n_{total} \sum_{i=1}^{k} n_{ii} - \sum_{i=j=1}^{k} (n_{i+} \times n_{+j})}{{n_{total}}^2 - \sum_{i=j=1}^{k} (n_{i+} \times n_{+j})} \qquad (4.8)$$

where i and j refer to row and column numbers of an error matrix, respectively. n is the number of samples assigned to one of k classes in the matrix. Khat > 0.8 represents strong agreement, 0.4 ≤Khat≤0.8 represents moderate agreement, Khat < 0.4 represents poor agreement (Landis and Koch, 1977).

Khat measures how well the classification result agrees with the reference data for a single error matrix. While the significance of the difference in accuracy, Z, allows for a statistical comparison between two error matrices. With this test, it is possible to determine if two error matrices are significantly different. In another word, it provides a means for statistically comparing the performance of two classifiers. The significance value with two independent Khat values can be evaluated in terms of the normal curve deviate as (Congalton and Green, 2009)

$$Z = \frac{|\hat{K}_1 - \hat{K}_2|}{\sqrt{\hat{var}(\hat{K}_1) + \hat{var}(\hat{K}_2)}} \tag{4.9}$$

and it can also test the significance of a single error matrix by:

$$Z = \frac{\hat{K}_1}{\sqrt{\hat{var}(\hat{K}_1)}} \tag{4.10}$$

where $\hat{K}_1$ and $\hat{K}_2$ denote independent Kappa coefficients of error matrix #1 and #2, respectively, Z is standardized and normally distributed, $\hat{var}(\hat{K}_1)$ and $\hat{var}(\hat{K}_2)$ denote

the corresponding approximate large sample variance which are computed using the Delta method as (Congalton and Green, 2009)

$$\widehat{var}(\hat{K}) = \frac{1}{n_{total}}\left\{\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2-\theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_2)^2(\theta_4-4\theta_2^2)}{(1-\theta_2)^4}\right\} \quad (4.11)$$

where

$$\theta_1 = \frac{\sum_{i=1}^{k} n_{ii}}{n_{total}} \quad (4.12)$$

$$\theta_2 = \frac{\sum_{i=j=1}^{k} n_{i+}n_{+j}}{{n_{total}}^2} \quad (4.13)$$

$$\theta_3 = \frac{\sum_{i=j=1}^{k} n_{ii}\left(n_{i+}+n_{+j}\right)}{{n_{total}}^2} \quad (4.14)$$

and

$$\theta_4 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} n_{ij}\left(n_{j+}+n_{+i}\right)^2}{{n_{total}}^3} \quad (4.15)$$

Despite the wide use and publicity of the approach represented by Equation (4.12) in remote sensing, many studies have inappropriately undertaken the significance analysis with related samples which, however, should be independent ones as the assumption of the approach states (Foody, 2004b). Most commonly in the remote sensing studies on accuracy comparison of different classifiers, like the present research, the same test samples are used to populate the error metrics from which the Khat values are derived. Therefore, an alternative technique, as described in (Chan et al., 2003; Watanachaturaporn et al., 2008), was used in this thesis to evaluate the significance of

difference for related samples. It is a non-parametric McNemar's test (Bradley, 1968; Agresti, 1996) based on the standardized normal statistic with a confidence interval of 95%:

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \tag{4.16}$$

where f_{ij} is the number of samples correctly classified by classifier i but misclassified by classifier j (see Table 4-5).

Table 4-3 The definition of matrix elements used in Equations (4.16) and (4.17)

Allocation	Classification 2		
	Correct	Incorrect	Σ
Classification 1			
Correct	f_{11}	f_{12}	
Incorrect	f_{21}	f_{22}	
Σ			

A χ^2distribution with 1 degree of freedom can be inferred from Equation (4.16) with the square of Z as (Agresti, 1996):

$$\chi^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \tag{4.17}$$

The statistical significance of difference is then tested by means of chi-squared distribution table with derived χ^2. The null hypothesis (H_0) of this two-sided test is that

there is no significant difference between the performances of two classifiers, and it would be rejected if $\chi^2 > 3.84$.

4.7 Chapter Summary

The methodologies from input data collection to output results evaluation have been stated in this chapter. First, the input features were generated from the LiDAR data and aerial imagery. Second, the training and reference samples were collected. Because training data selection was requisite as an initial step for further supervised feature selection and classification approaches, and reference data is to compare with the classification results. Third, different classification schemes based on MLC, SVM, and RF classifiers were performed with or without feature selection techniques including PCA and RF. Three different training sample sizes were also used to assess the effect of training data. Finally, both quantitative and qualitative evaluation methods were applied to the classification results.

Chapter 5
Results and Discussion

This chapter provides all the test results and discussions. Section 5.1 presents all the selected features from PCA based and RF based strategies. Section 5.2 reports the quantitative assessment results from four aspects: overall accuracy, kappa test, User's and Producer's accuracies, and Mcnemar test. The results from qualitative visual assessment are stated in Section 5.3, followed by a processing speed comparison in Section 5.4. Finally, Section 5.5 provides the summary of this chapter.

5.1 Feature Selection Results

The number of features generated using PCA method from image feature group is 8, from LiDAR feature group is 12, from all feature group is 20. The features selected with the RF methods are shown in Table 5-1. In this table, for each feature group with tr#250, tr#500, or tr#1000, the optimal features were selected by the RF based feature selection package 'varSelRF' with 20% iterative elimination. In most cases, RF based feature selection method provided lower number of features than PCA based method. Overall, the most frequent selected features were RGB, Eigen3, Fr-Lr, DSM, LiDAR-TVI, HDiff, and Contrast attributes from GLCM.

Table 5-1 Feature selection results from RF based method

	Image Feature Group			LiDAR Feature Group			All Feature Group		
	tr#250	tr#500	tr#1000	tr#250	tr#500	tr#1000	tr#250	tr#500	tr#1000
Red	√	√	√	×	×	×	×	√	×
Red_Contrast	×	×	×	×	×	×	×	×	×
Red_Entropy	×	×	×	×	×	×	×	×	×
Red_Correlation	√	√	√	×	×	×	×	√	×
Green	√	√	√	×	×	×	√	√	×
Green_Contrast	√	×	√	×	×	×	×	×	×
Green_Entropy	×	×	√	×	×	×	×	×	×
Green_Correlation	×	×	√	×	×	×	×	√	×
Blue	√	√	√	×	×	×	√	√	√
Blue_Contrast	×	×	√	×	×	×	×	√	×
Blue_Entropy	×	×	×	×	×	×	×	×	×
Blue_Correlation	×	×	×	×	×	×	×	×	×
DSM	×	×	×	√	√	√	√	√	√
DSM_Contrast	×	×	×	×	×	×	×	√	×
DSM_Entropy	×	×	×	×	×	×	×	√	×
DSM_Correlation	×	×	×	√	√	√	√	√	√
Intensity	×	×	×	√	√	√	√	√	√
Intensity_Contrast	×	×	×	×	×	×	×	√	×
Intensity_Entropy	×	×	×	×	×	×	×	×	×
Intensity_Correlation	×	×	×	×	×	×	×	×	×
Fr-Lr	×	×	×	√	√	√	√	√	√
HVar	×	×	×	×	√	×	√	√	√
HDiff	×	×	×	√	√	√	√	√	√
Eigen1	×	×	×	×	×	×	√	√	×
Eigen2	×	×	×	×	×	×	×	×	×
Eigen3	×	×	×	×	√	√	√	√	√
Anisotropy	×	×	×	×	×	×	×	×	×
Planarity	×	×	×	×	×	×	√	√	√
Linearity	×	×	×	×	×	×	√	√	×
Sphericity	×	×	×	×	×	×	×	×	×
LiDAR-TVI	×	×	×	×	×	×	√	√	√

Note: for each column, '√' indicates the selected feature, '×' indicates the excluded feature.

5.2 Classification Accuracies

5.2.1 Overall Accuracy Results

Overall classification accuracies of each classifier with different training data sizes and feature groups are reported in Tables 5-2 to 5-4. These values indicate the percentage of total number of pixels correctly classified out of 2500 reference pixels.

Table 5-2 shows the overall accuracies from image feature group classification. The highest overall accuracy in this table is 65.44% achieved by RF-SVM with 4000 training samples. Overall, SVM based and RF based classifiers outperformed than MLC. RF based feature selection achieved better classification than PCA based feature selection in most cases. On average, Overall accuracies derived from classification with feature selection were higher than those without feature selection. However, all these classification results in terms of overall accuracy were not good enough for land cover classification, considering the accuracy values only ranged from 28.76% to 65.44%.

Table 5-3 lists the classification overall accuracies using LiDAR features. In this table, the highest overall classification accuracy (82.32%) was obtained with 4000 training samples by RF classifier, followed by SVM and RF-RF with 4000 training samples (81.72%). The lowest accuracy in this table is 73.60% achieved by RF-MLC with 4000 samples. All experiment accuracies from LiDAR feature classification were much greater than those from image feature classification. The overall accuracy improved

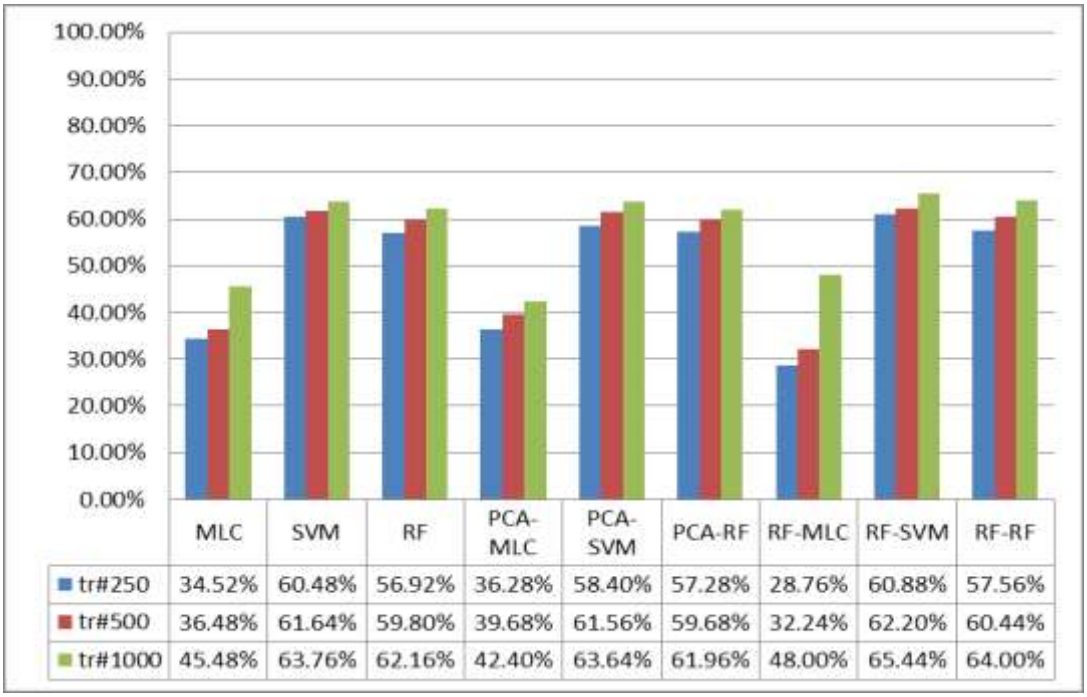
by approximately 0.2% - 4% when using RF based feature selection method than PCA based method in most cases. In this table, all accuracies derived from SVM based and RF based classifications were higher than those from MLC based classification.

Table 5-4 presents the classification overall accuracies based on all feature group. The lowest classification accuracy is 78.48% obtained by MLC with 1000 training samples. In contrast, RF with 4000 training samples achieved highest overall accuracy, 88.00%. It demonstrates that all the supervised classification schemes in this study performed accurately and achieved satisfied overall accuracy by fusion of LiDAR data and aerial imagery. SVM and RF classifiers improved the overall accuracy by about 3% - 5% than MLC classifiers. RF based feature selection resulted in up to 3.5% greater accuracy than PCA based feature selection.

Generally, MLC, PCA-MLC, and RF-MLC were less accurate than other classifiers. RF and SVM gave significantly higher accuracies than MLC in all three level of training size. The overall accuracy results clearly showed the positive impact of the use of advanced machine learning algorithms like SVM and RF. The classifications after two feature selection methods were also compared in terms of overall accuracy. In general, RF-RFs always showed higher overall accuracies in comparison with PCA-RFs. Most RF-SVMs and RF-MLCs performed more accurate than PCA-SVMs and PCA-MLCs, respectively. Therefore, on average, RF based feature selection methods offered more accurate classification results than PCA based feature selection methods in this study.

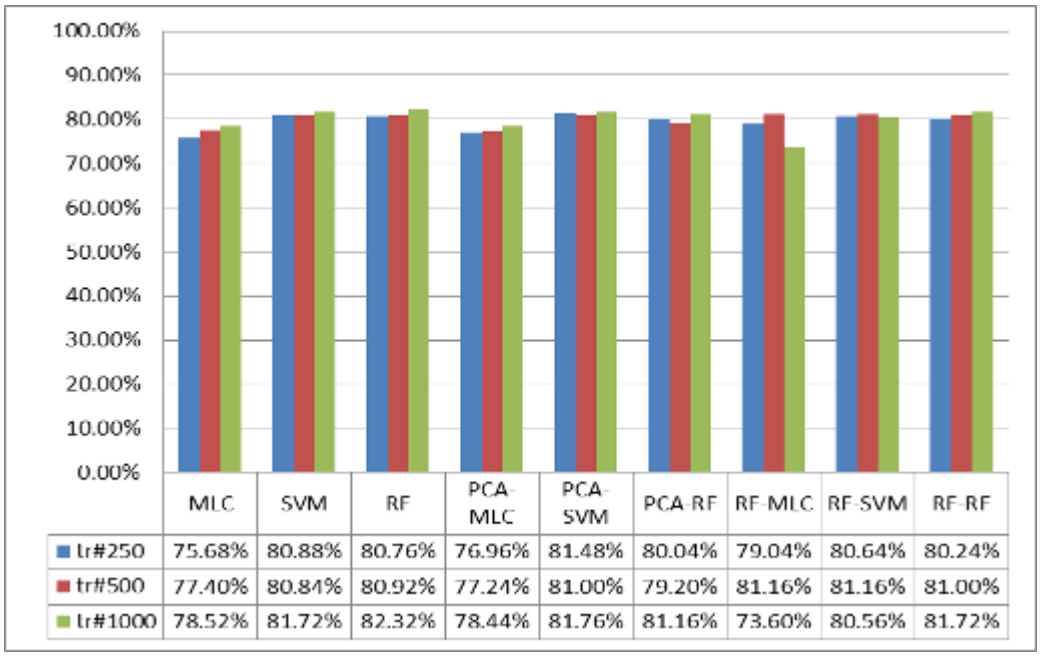
One of the goals of this experiment was to assess the effectiveness of the proposed classification schemes for a relative small training set (tr#250, tr#500, tr#1000) and the influence of different training sizes. Reorganizing the numbers in Tables 5-2 to 5-4 shows that the impact of training data size on algorithm overall accuracies. As expected, increases in training data size generally led to improved performances in most experiments. However, SVM based and RF based classification schemes performed more stably over different training sizes.

Table 5-2 Classification overall accuracies using image feature group



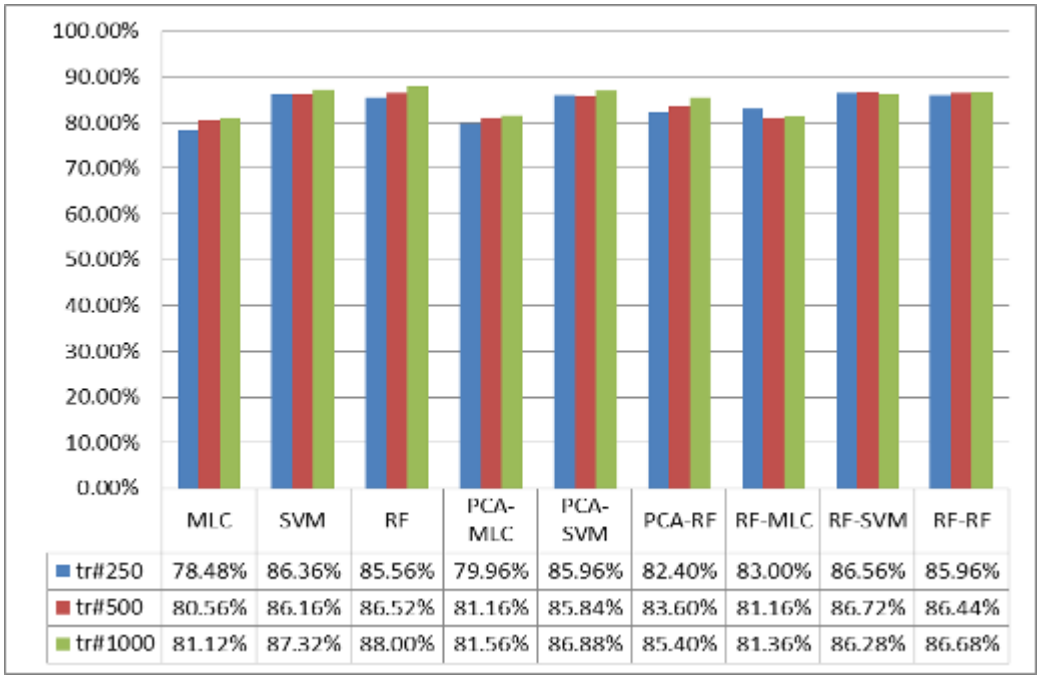
	MLC	SVM	RF	PCA-MLC	PCA-SVM	PCA-RF	RF-MLC	RF-SVM	RF-RF
tr#250	34.52%	60.48%	56.92%	36.28%	58.40%	57.28%	28.76%	60.88%	57.56%
tr#500	36.48%	61.64%	59.80%	39.68%	61.56%	59.68%	32.24%	62.20%	60.44%
tr#1000	45.48%	63.76%	62.16%	42.40%	63.64%	61.96%	48.00%	65.44%	64.00%

Table 5-3 Classification overall accuracies using LiDAR feature group



	MLC	SVM	RF	PCA-MLC	PCA-SVM	PCA-RF	RF-MLC	RF-SVM	RF-RF
tr#250	75.68%	80.88%	80.76%	76.96%	81.48%	80.04%	79.04%	80.64%	80.24%
tr#500	77.40%	80.84%	80.92%	77.24%	81.00%	79.20%	81.16%	81.16%	81.00%
tr#1000	78.52%	81.72%	82.32%	78.44%	81.76%	81.16%	73.60%	80.56%	81.72%

Table 5-4 Classification overall accuracies using all feature group



	MLC	SVM	RF	PCA-MLC	PCA-SVM	PCA-RF	RF-MLC	RF-SVM	RF-RF
tr#250	78.48%	86.36%	85.56%	79.96%	85.96%	82.40%	83.00%	86.56%	85.96%
tr#500	80.56%	86.16%	86.52%	81.16%	85.84%	83.60%	81.16%	86.72%	86.44%
tr#1000	81.12%	87.32%	88.00%	81.56%	86.88%	85.40%	81.36%	86.28%	86.68%

5.2.2 Kappa Test Results

As explained in Section 4.5, the Khat values denote a measure of agreement between the remotely sensed classification and the reference data. Table 5-5 presents the results of the Kappa analysis on the individual error matrices derived from each classification scheme on each training size level. According to Landis and Koch (1977), all the MLC based classification schemes, including MLC, PCA-MLC, and RF-MLC, showed poor agreement in image feature group (see Table 5-5a). All of the Khat values derived from LiDAR feature group classification results showed moderate agreement (see Table 5-5b); however several SVM based and RF based classification schemes using all feature group showed strong agreement (see Table 5-5c). On average, the Khat values derived from RF and RF-RF were greater than other classifiers. Also in Table 5-5, the Z statistic discussed in Chapter 3 is provided. At the 95% confidence level, if the value of the Z statistic is greater than 1.96, the classification is significantly better than a random classification (Congalton and Green, 2009). The Z statistic values for all the supervised classifiers in Table 5-5 were more than 9, therefore, all the classifications were significantly better than random results.

Table 5-5 Individual error matrix kappa analysis results

(a) Kappa Analysis based on image feature group

	Kappa Analysis Parameters	MLC	SVM	RF	PCA-MLC	PCA-SVM	PCA-RF	RF-MLC	RF-SVM	RF-RF
tr#250	Khat	0.15	0.42	0.37	0.16	0.40	0.37	0.09	0.42	0.38
	Z Statistic	14.14	30.08	26.89	15.10	29.17	26.77	9.47	30.35	26.86
tr#500	Khat	0.18	0.44	0.41	0.21	0.43	0.41	0.13	0.43	0.42
	Z Statistic	16.65	31.84	29.40	19.51	31.32	29.40	15.33	31.28	29.80
tr#1000	Khat	0.26	0.46	0.44	0.28	0.46	0.43	0.27	0.48	0.47
	Z Statistic	21.68	33.46	30.99	24.15	32.83	31.35	23.99	35.15	33.37

Note: the Khat value highlighted by a box indicates poor agreement.

(b) Kappa Analysis based on LiDAR feature group

	Kappa Analysis Parameters	MLC	SVM	RF	PCA-MLC	PCA-SVM	PCA-RF	RF-MLC	RF-SVM	RF-RF
tr#250	Khat	0.63	0.72	0.72	0.66	0.72	0.71	0.69	0.71	0.71
	Z Statistic	50.21	62.79	62.84	54.81	63.96	61.21	58.05	62.06	61.59
tr#500	Khat	0.66	0.72	0.72	0.67	0.72	0.69	0.72	0.72	0.72
	Z Statistic	53.45	62.83	63.13	56.22	63.12	59.85	63.74	63.98	63.33
tr#1000	Khat	0.68	0.73	0.74	0.68	0.73	0.72	0.62	0.71	0.73
	Z Statistic	55.86	64.88	66.05	57.24	64.92	64.05	49.94	62.40	64.91

(c) Kappa Analysis based on all feature group

	Kappa Analysis Parameters	MLC	SVM	RF	PCA-MLC	PCA-SVM	PCA-RF	RF-MLC	RF-SVM	RF-RF
tr#250	Khat	0.68	0.79	0.78	0.70	0.79	0.74	0.75	**0.80**	0.79
	Z Statistic	56.78	77.02	75.04	60.79	75.80	66.64	67.10	77.60	76.49
tr#500	Khat	0.71	0.79	**0.80**	0.72	0.79	0.75	0.72	**0.80**	**0.80**
	Z Statistic	60.91	76.83	78.19	63.24	75.54	69.06	63.74	78.38	77.76
tr#1000	Khat	0.72	**0.81**	**0.82**	0.73	**0.80**	0.78	0.72	0.79	**0.80**
	Z Statistic	62.78	80.76	83.76	64.19	79.07	74.05	62.86	77.12	78.54

Note: the Khat value presented boldface shows strong agreement.

5.2.3 User's and Producer's Accuracy Results

The User's and Producer's accuracies of individual land cover type were listed in Tables 5-6 to 5-8. Table 5-6 provides User's and Producer's accuracies derived from different classification scheme based on image feature group. Most of the Producer's accuracies of four land cover types were lower than 70%, the lowest value is only 2.83%, which were not adequate for land cover mapping. On all training sample sizes, MLC based classification schemes, including MLC, PCA-MLC, and RF-MLC, presented much stronger imbalance between different classes than other classification schemes.

In Table 5-7, class-specific classification accuracies using LiDAR feature group are listed. The Producer's accuracies of 'building' were ranged from 71.40% to 89.82%. The 'tree' class attained the Producer's accuracies ranged from 76.08% to 90.84%. The Producer's accuracies of 'grass' were ranged from 60.85% to 85.38%. And the Producer's accuracies of 'road' were ranged from 65.66% to 80.45%. On average, SVM classifiers improved the Producer's accuracy of 'building' by about 2% - 10% compared to MLC classifiers; RF classifiers improved by about 2% - 12% compared to MLC classifiers.

In Table 5-8, the User's and Producer's accuracies of classifications using all feature group are listed. The highest Producer's accuracy of the "building" class was obtained by RF-MLC, SVM, and RF with tr#1000 (89.12%). The three highest Producer's accuracies of "tree" were derived from MLC with tr#250 (92.88%) and PCA-

MLC with tr#250 (92.62%), followed by RF with tr#500 and tr#1000 (92.37%). The highest Producer's accuracy of "grass" was generated by RF-SVM with tr#1000 (90.09%). The classification schemes performed relatively weaker on the "road" land cover type resulting in Producer's accuracies between 74.64% (MLC with tr#250) and 86.34% (RF-SVM with tr#250). These values are lower than those obtained for the other classes. SVM based classifications, including SVM, PCA-SVM, RF-SVM, were the best on classification of "building" and "road", resulting in Producer's accuracies between 84.83% and 89.12%. In contrast, RF based classifications, including RF, PCA-RF, RF-RF, provided the highest Producer's accuracies of "tree" and "grass" classes from 84.91% to 92.37%. Therefore, SVM based classifications performed well for impervious surface while RF based classifications provided better results on pervious classes based on 27 experiments with all feature group.

On average, the Producer's accuracies achieved by SVM based and RF based classification schemes are much higher than those obtained by MLC based ones. Overall, SVM based and RF based classification schemes produced more balanced Producer's and User's accuracies than MLC based ones. When using MLC classifier to evaluate the feature selection methods, the RF based feature selection generated more improvement than PCA based feature selection method in terms of the Producer's accuracy of each land cover type. In another words, RF-MLCs generated higher Producer's accuracy than PCA-MLCs on each land cover type. Both RF and PCA based feature selection methods improved the MLC classification results.

Table 5-6 Class-specific classification accuracies using image feature group

Class-specific classification accuracies using image feature group and tr#250 (%)							Class-specific classification accuracies using image feature group and tr#500 (%)						Class-specific classification accuracies using image feature group and tr#1000 (%)					
	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
	MLC		SVM		RF		MLC		SVM		RF		MLC		SVM		RF	
Building	35.44	25.28	58.42	42.47	52.46	39.19	61.93	27.51	61.58	47.56	54.91	42.70	54.04	33.73	59.12	50.07	55.09	44.41
Tree	88.04	27.64	60.31	55.50	58.02	52.66	81.42	36.12	58.52	51.80	59.80	52.93	82.70	34.39	59.03	54.21	61.58	56.67
Grass	11.79	33.33	61.32	43.92	63.21	38.40	20.75	52.38	66.98	43.29	62.26	41.51	13.68	34.52	69.81	44.18	65.09	46.31
Road	21.89	77.54	61.28	81.77	57.51	79.79	14.72	78.95	61.74	82.63	61.51	81.09	35.85	85.13	66.19	82.42	64.91	80.52
	PCA-MLC		PCA-SVM		PCA-RF		PCA-MLC		PCA-SVM		PCA-RF		PCA-MLC		PCA-SVM		PCA-RF	
Building	48.25	25.56	58.77	41.72	51.58	41.12	81.75	28.93	57.72	45.13	57.19	42.84	83.51	30.16	56.32	49.38	59.12	46.81
Tree	83.46	31.54	58.02	52.53	49.11	49.87	68.70	47.20	58.27	55.31	54.96	52.94	65.90	49.43	55.47	56.92	52.16	52.03
Grass	6.60	48.28	64.15	39.42	59.43	34.05	24.53	72.22	62.74	40.30	63.68	40.79	16.98	52.17	67.45	40.51	63.68	41.80
Road	21.89	81.69	57.43	82.90	61.81	79.67	15.40	83.27	64.00	82.57	61.51	81.50	21.81	87.84	68.60	81.60	65.81	82.03
	RF-MLC		RF-SVM		RF-RF		RF-MLC		RF-SVM		RF-RF		RF-MLC		RF-SVM		RF-RF	
Building	43.16	20.87	57.72	45.13	57.02	40.98	83.16	26.66	58.25	46.30	57.54	44.63	31.58	42.06	63.51	51.13	60.00	46.72
Tree	86.01	30.42	46.82	51.40	50.89	50.13	69.21	43.45	41.98	51.89	51.15	51.94	89.06	28.32	58.02	57.58	62.09	60.10
Grass	2.83	25.00	68.40	40.62	54.25	37.83	3.77	33.33	71.23	42.30	65.09	40.95	9.91	24.42	66.04	45.16	67.92	47.21
Road	9.74	69.35	65.21	81.82	60.30	79.58	3.92	72.22	68.45	81.86	63.70	81.08	48.98	86.53	68.38	83.43	65.66	82.31

Note: PA: Producer's Accuracy, UA: User's Accuracy.

Table 5-7 Class-specific classification accuracies using LiDAR feature group

Class-specific classification accuracies using LiDAR feature group and tr#250 (%)							Class-specific classification accuracies using LiDAR feature group and tr#500 (%)						Class-specific classification accuracies using LiDAR feature group and tr#1000 (%)					
	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
	MLC		SVM		RF		MLC		SVM		RF		MLC		SVM		RF	
Building	71.40	71.65	81.93	86.00	83.33	81.90	75.79	71.05	83.68	81.82	83.68	82.67	82.81	65.46	85.09	83.19	84.56	83.83
Tree	83.72	65.02	84.99	75.06	86.26	71.82	82.19	70.68	82.95	74.09	86.01	73.16	79.90	73.36	84.73	77.44	85.50	75.17
Grass	60.85	47.25	84.91	44.44	82.55	47.43	70.28	50.51	82.08	45.79	82.08	46.40	70.28	60.82	83.02	45.71	82.08	48.60
Road	77.51	89.07	78.57	94.04	77.74	95.46	77.81	90.44	78.79	95.17	78.04	95.21	77.58	92.95	79.17	95.19	80.45	95.18
	PCA-MLC		PCA-SVM		PCA-RF		PCA-MLC		PCA-SVM		PCA-RF		PCA-MLC		PCA-SVM		PCA-RF	
Building	78.60	72.14	84.91	85.06	83.33	80.92	85.44	68.21	83.68	83.10	82.63	80.65	86.67	65.87	85.96	82.49	84.74	81.59
Tree	88.30	63.67	83.46	77.36	85.75	74.07	85.75	70.35	83.72	74.77	84.99	72.93	83.46	72.41	82.70	78.88	84.48	75.63
Grass	71.23	50.67	82.55	44.76	79.72	44.13	76.89	49.54	82.08	45.31	83.02	43.24	70.75	57.03	82.55	45.22	84.43	45.66
Road	73.81	94.40	79.25	94.09	76.98	94.88	71.25	96.52	78.87	94.83	75.40	95.05	74.64	95.65	79.55	95.21	78.11	96.10
	RF-MLC		RF-SVM		RF-RF		RF-MLC		RF-SVM		RF-RF		RF-MLC		RF-SVM		RF-RF	
Building	84.74	68.90	82.46	80.76	82.81	82.81	86.14	70.95	83.86	84.60	82.98	83.72	89.82	55.90	83.33	82.61	83.16	83.01
Tree	80.15	73.26	78.37	77.78	82.44	69.53	90.84	75.16	83.97	76.39	86.26	71.22	76.08	77.06	82.95	74.94	84.99	73.41
Grass	72.17	53.13	84.91	44.78	83.96	47.09	85.38	59.34	85.38	43.83	82.55	47.30	75.00	54.83	84.43	44.53	83.96	48.24
Road	77.36	94.82	79.85	94.46	77.89	95.03	75.47	97.28	78.49	95.41	78.34	95.32	65.66	96.03	78.04	95.04	79.77	95.66

Note: PA: Producer's Accuracy, UA: User's Accuracy.

Table 5-8 Class-specific classification accuracies using all feature group

Class-specific classification accuracies using all feature group and tr#250 (%)							Class-specific classification accuracies using all feature group and tr#500 (%)						Class-specific classification accuracies using all feature group and tr#1000 (%)					
	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
	MLC		SVM		RF		MLC		SVM		RF		MLC		SVM		RF	
Building	75.96	70.29	86.67	84.73	85.61	84.28	81.75	69.66	87.54	83.72	87.37	85.13	88.95	66.71	89.12	85.38	89.12	86.69
Tree	92.88	69.13	87.79	85.61	92.11	76.69	90.08	75.00	87.79	82.14	92.37	79.61	88.30	78.33	88.55	85.50	92.37	82.31
Grass	82.55	59.32	87.74	57.23	87.74	60.78	82.55	62.50	87.74	58.49	88.21	61.51	80.19	66.41	87.74	59.81	89.15	64.07
Road	74.64	93.21	85.58	95.37	83.25	96.50	76.91	94.44	84.83	96.40	84.15	96.54	75.77	96.45	86.11	96.12	86.04	96.77
	PCA-MLC		PCA-SVM		PCA-RF		PCA-MLC		PCA-SVM		PCA-RF		PCA-MLC		PCA-SVM		PCA-RF	
Building	81.05	70.64	86.49	84.42	84.21	78.43	86.84	69.04	86.32	84.68	84.91	78.96	87.89	69.20	88.25	84.68	85.61	80.79
Tree	92.62	67.78	87.28	85.75	88.04	75.55	91.09	74.27	86.77	82.57	89.31	77.48	89.06	75.76	87.53	86.00	88.30	80.70
Grass	81.13	63.00	87.74	55.52	87.26	53.62	82.55	65.54	87.26	56.40	84.91	57.32	82.55	64.58	87.74	58.31	87.74	60.59
Road	75.55	96.62	85.06	95.43	79.17	96.68	75.55	96.81	85.13	95.76	81.13	95.98	76.45	97.12	85.96	95.96	84.08	96.12
	RF-MLC		RF-SVM		RF-RF		RF-MLC		RF-SVM		RF-RF		RF-MLC		RF-SVM		RF-RF	
Building	83.16	73.37	85.79	85.94	85.44	85.74	86.14	70.95	87.89	85.79	86.49	84.56	89.12	64.88	87.37	84.69	87.54	84.72
Tree	90.08	74.53	88.04	82.58	91.60	76.43	90.84	75.16	89.82	82.09	92.11	78.70	78.12	79.33	87.53	82.10	89.57	80.18
Grass	84.43	66.30	87.26	59.29	88.68	60.65	85.38	59.34	86.32	60.20	87.74	62.84	83.02	70.12	90.09	59.13	88.68	62.25
Road	80.60	96.30	86.34	95.33	84.08	96.79	75.47	97.28	85.36	95.69	84.53	96.47	78.72	96.66	84.83	96.07	85.13	96.41

Note: PA: Producer's Accuracy, UA: User's Accuracy.

5.2.4 McNemar Test Comparison

The classification results derived from different classification schemes were compared using the McNemar test with a 95% confidence interval in Tables 5-9 to 5-11. Table 5-9 presents the χ^2 values derived from McNemar test on classification using image feature group. If χ^2value is larger than 3.84, two classifiers are significantly different. Table 5-10 lists the McNemar test results derived from classification comparison based LiDAR feature group. In Table 5-11, the χ^2 values of McNemar test on classifications using all feature group are stated.

In general, the McNemar test results showed that all the MLC classifiers provide significantly different results against the SVM and RF classifers. The differences of classification accuracies between SVM and RF were not statistically significant, since most χ^2 values from McNemar's test were less than 3.84. Thus, both SVM and RF perform equally well. Of the two, it is better to choose the cheaper, quicker, or more efficient classification scheme. After comparing classifiers with and without feature selection methods, it described that significant differences were occurred more with relatively larger training sample size (tr#1000) than smaller one (tr#250), especially for SVM and RF. All PCA-RFs showed significant differences against RF-RFs in all feature group. Overall, most PCA-MLCs and RF-SVMs were significantly different. While almost half PCA-SVMs derived significantly different results compared with RF-SVMs. It implied that it is worth choosing one of the feature selection methods with higher classification accuracy.

Table 5-9 McNemar test results with image feature group

	tr#250		tr#500		tr#1000	
	χ^2	Significant?	χ^2	Significant?	χ^2	Significant?
MLC vs. SVM	359.69	Yes	339.61	Yes	220.54	Yes
MLC vs. RF	279.50	Yes	307.04	Yes	190.04	Yes
MLC vs. RF-MLC	41.64	Yes	23.12	Yes	9.00	Yes
MLC vs. PCA-MLC	5.98	Yes	20.78	Yes	10.68	Yes
RF-MLC vs. PCA-MLC	65.21	Yes	71.48	Yes	22.69	Yes
SVM vs. RF	13.73	Yes	3.74	No	3.64	No
SVM vs. RF-SVM	0.18	No	0.33	No	5.92	Yes
SVM vs. PCA-SVM	9.59	Yes	0.01	No	0.03	No
RF-SVM vs. PCA-SVM	5.90	Yes	0.49	No	6.47	Yes
RF vs. RF-RF	0.58	No	0.44	No	8.08	Yes
RF vs. PCA-RF	0.14	No	0.02	No	0.05	No
RF-RF vs. PCA-RF	0.08	No	0.51	No	5.23	Yes

Table 5-10 McNemar test results with LiDAR feature group

	tr#250		tr#500		tr#1000	
	χ^2	Significant?	χ^2	Significant?	χ^2	Significant?
MLC vs. SVM	44.24	Yes	24.33	Yes	20.38	Yes
MLC vs. RF	45.43	Yes	26.16	Yes	29.21	Yes
MLC vs. RF-MLC	20.63	Yes	3.24	No	43.60	Yes
MLC vs. PCA-MLC	3.94	Yes	0.06	No	0.03	No
RF-MLC vs. PCA-MLC	11.96	Yes	6.22	Yes	48.97	Yes
SVM vs. RF	0.06	No	0.03	No	1.77	No
SVM vs. RF-SVM	0.23	No	0.39	No	5.88	Yes
SVM vs. PCA-SVM	2.47	No	0.24	No	0.01	No
RF-SVM vs. PCA-SVM	2.44	No	0.10	No	6.16	Yes
RF vs. RF-RF	1.04	No	0.03	No	1.92	No
RF vs. PCA-RF	1.80	No	10.69	Yes	6.05	Yes
RF-RF vs. PCA-RF	0.10	No	9.42	Yes	0.97	No

Table 5-11 McNemar test results with all feature group

	tr#250		tr#500		tr#1000	
	χ^2	Significant?	χ^2	Significant?	χ^2	Significant?
MLC vs. SVM	112.49	Yes	73.68	Yes	86.73	Yes
MLC vs. RF	105.48	Yes	83.78	Yes	111.22	Yes
MLC vs. RF-MLC	57.26	Yes	1.20	No	0.17	No
MLC vs. PCA-MLC	7.73	Yes	1.35	No	0.92	No
RF-MLC vs. PCA-MLC	30.72	Yes	0.00	No	0.15	No
SVM vs. RF	2.47	No	0.57	No	2.39	No
SVM vs. RF-SVM	0.27	No	2.45	No	5.73	Yes
SVM vs. PCA-SVM	2.00	No	0.97	No	2.20	No
RF-SVM vs. PCA-SVM	2.32	No	5.38	Yes	2.10	No
RF vs. RF-RF	1.19	No	0.10	No	14.73	Yes
RF vs. PCA-RF	31.36	Yes	30.45	Yes	27.98	Yes
RF-RF vs. PCA-RF	35.84	Yes	27.85	Yes	4.70	Yes

5.3 Visual Assessment

This section is based on visual comparison of information content of the generated land cover maps. It is more a qualitative rather than a quantitative analysis in previous section. To illustrate the classification results, a series of land cover maps are presented in Figures 5.1 to 5.9. Figure 5.1 shows the classified land cover images achieved by MLC using three feature groups (image, LiDAR, and all) and three training sample sizes (tr#250, tr#500, tr#1000). Figure 5.2 illustrates the classification results by SVM. Figure 5.3 shows the classification results using RF. Figure 5.4 presents the classified image by PCA-MLC. Figure 5.5 presents the classification result images by RF-MLC. Figure 5.6 illustrates the classification result images by PCA-SVM. Figure 5.7 shows the classified land cover images by RF-SVM. Figure 5.8 shows the classification result images by

PCA-RF. Figure 5.9 shows the classification result images by RF-RF. The original aerial imagery has shown below each figure to compare with the classification results.

Generally speaking, the visual assessment supports the results of the statistical accuracy assessment in previous section. Image feature group generated most noisy maps. And the maps from LiDAR feature group showed the general structures of the study area, however, appear many errors on the edges along different land cover objects. This drawback is significantly reduced by the all feature group. Almost all areas can be assigned properly to a specific class using all feature group. Borders between different classes appear clearer and are easier to identify. However, some errors are still inherent, such as confusion between “building” and “tree”, due to the similarity between these classes within the feature space. The visible comparison of the classification result maps underlines the statistical assessment: the maps produced by SVM based and RF based classifications are more homogeneous over each land cover type than MLC based ones. Of all, all feature group result in the least noisy maps.

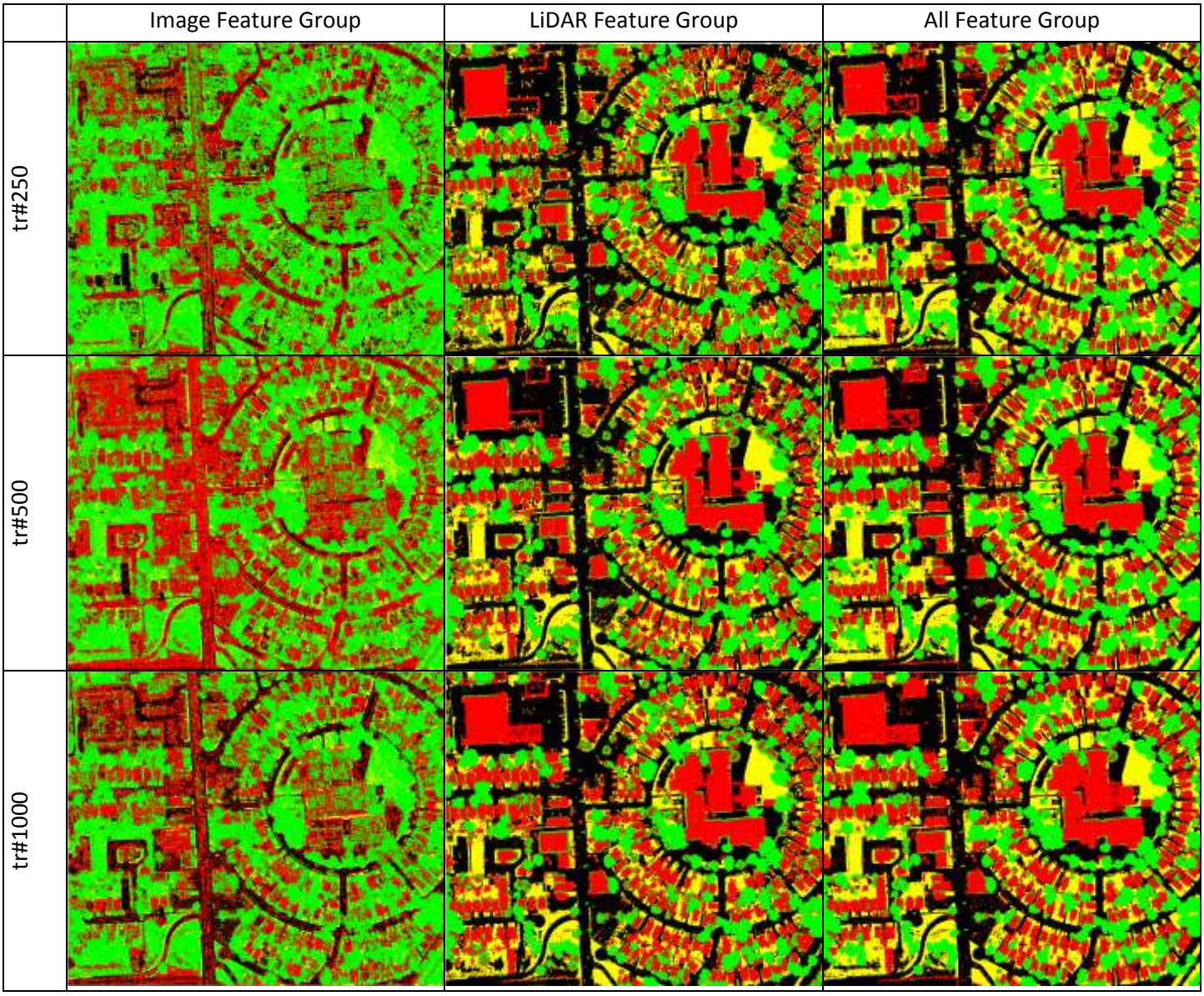


Figure 5.1 Classification maps achieved by MLC

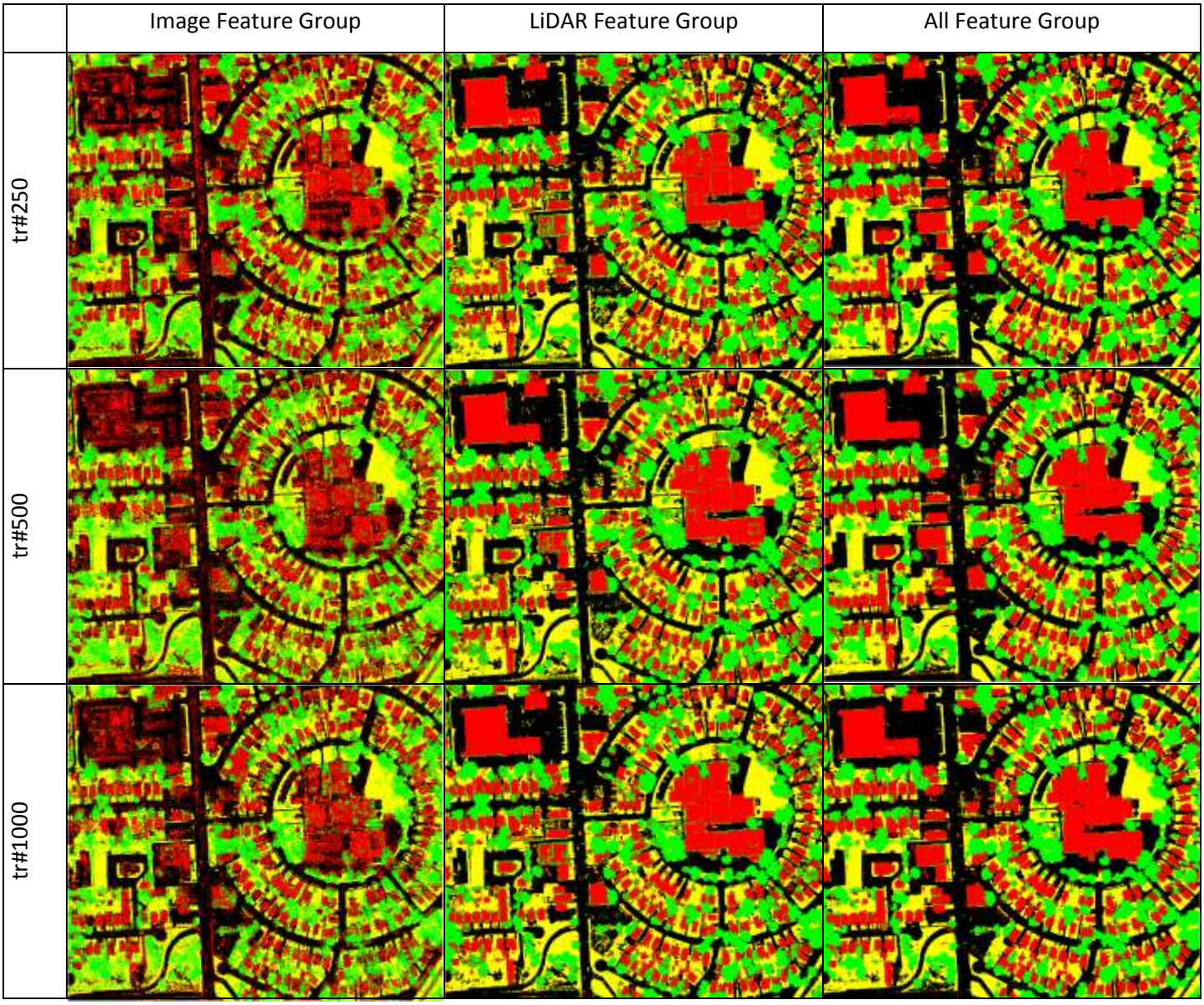
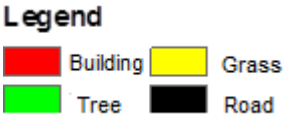


Figure 5.2 Classification maps achieved by SVM



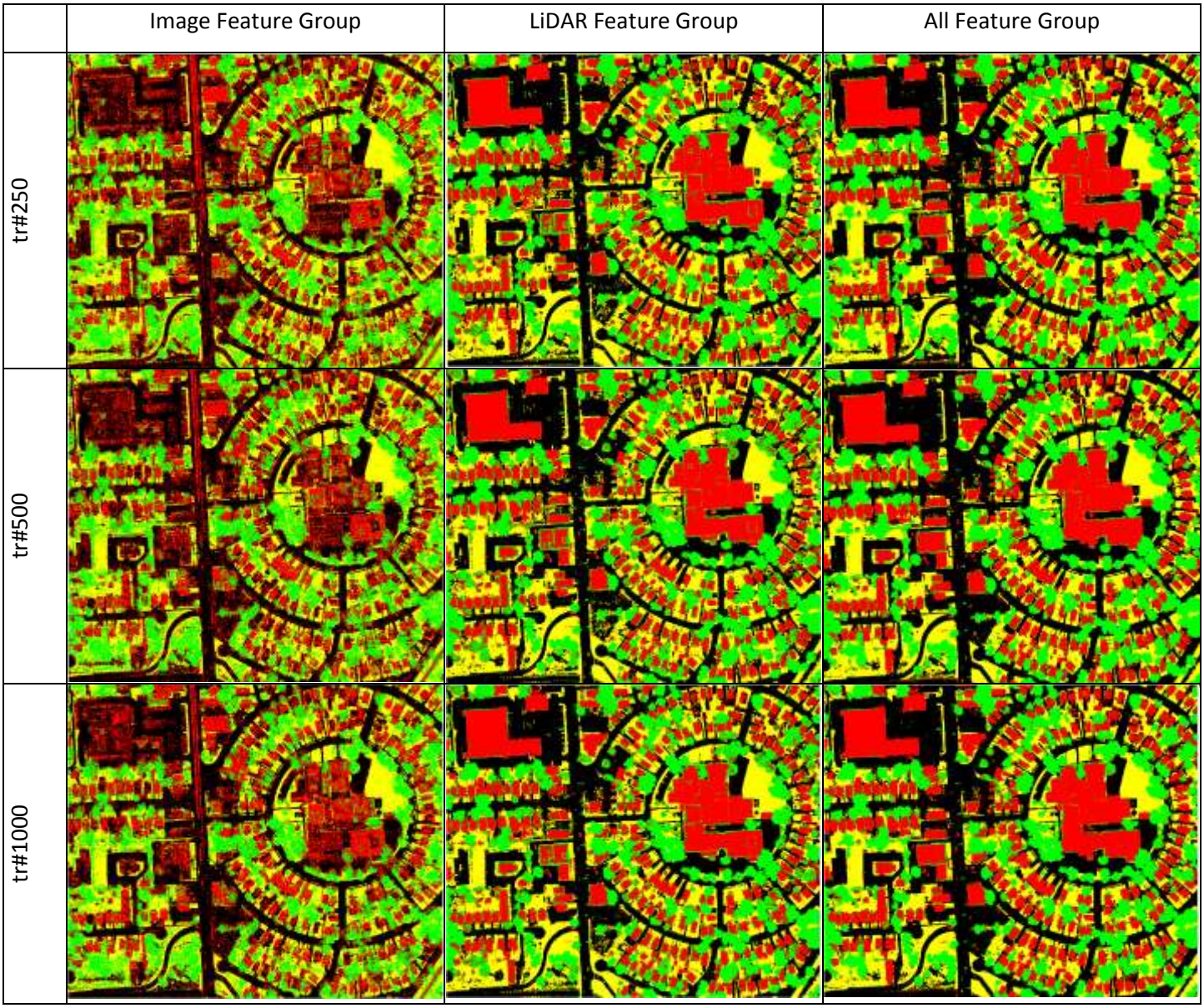
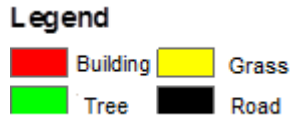


Figure 5.3 Classification maps achieved by RF



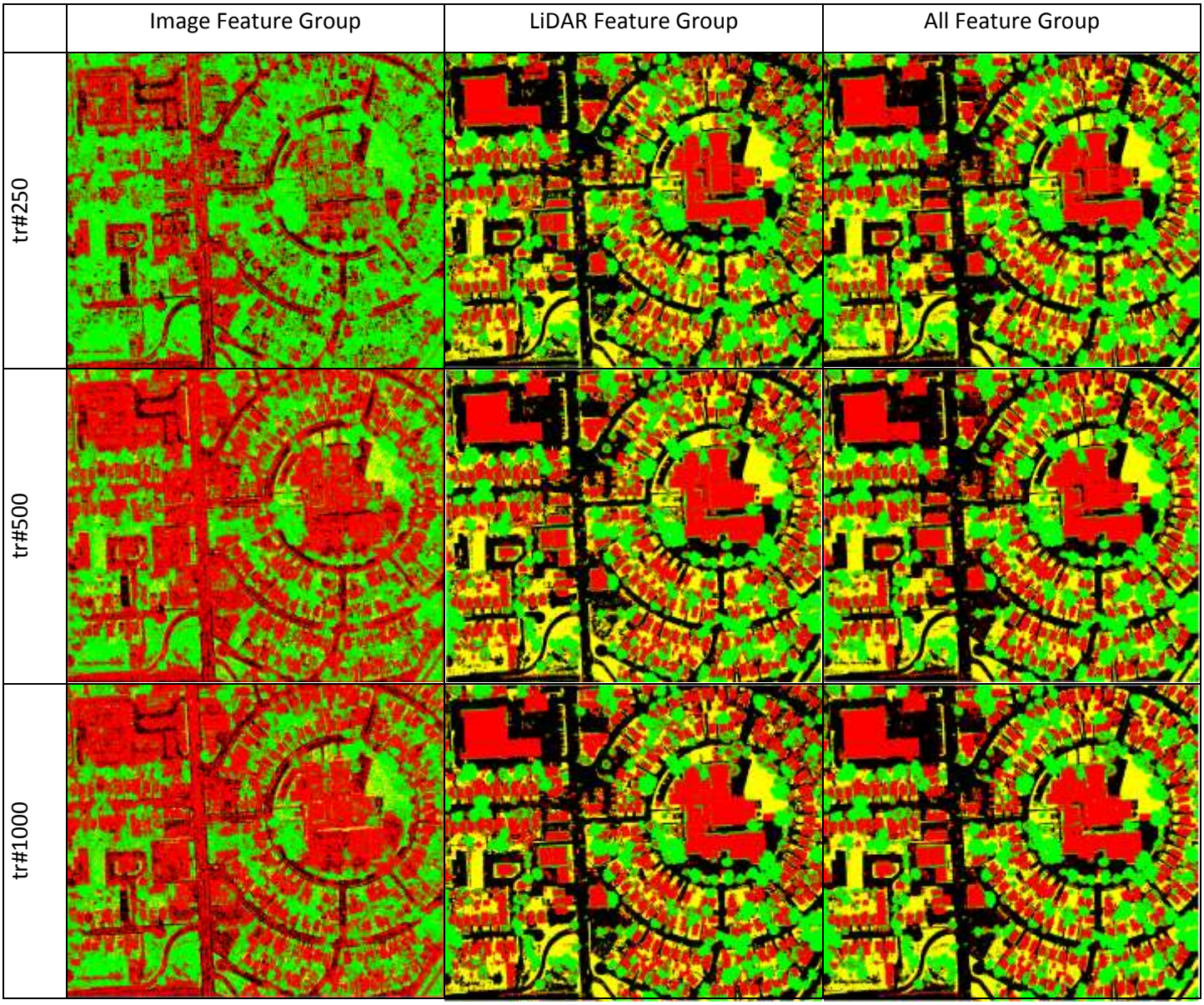
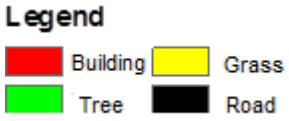


Figure 5.4 Classification maps achieved by PCA-MLC



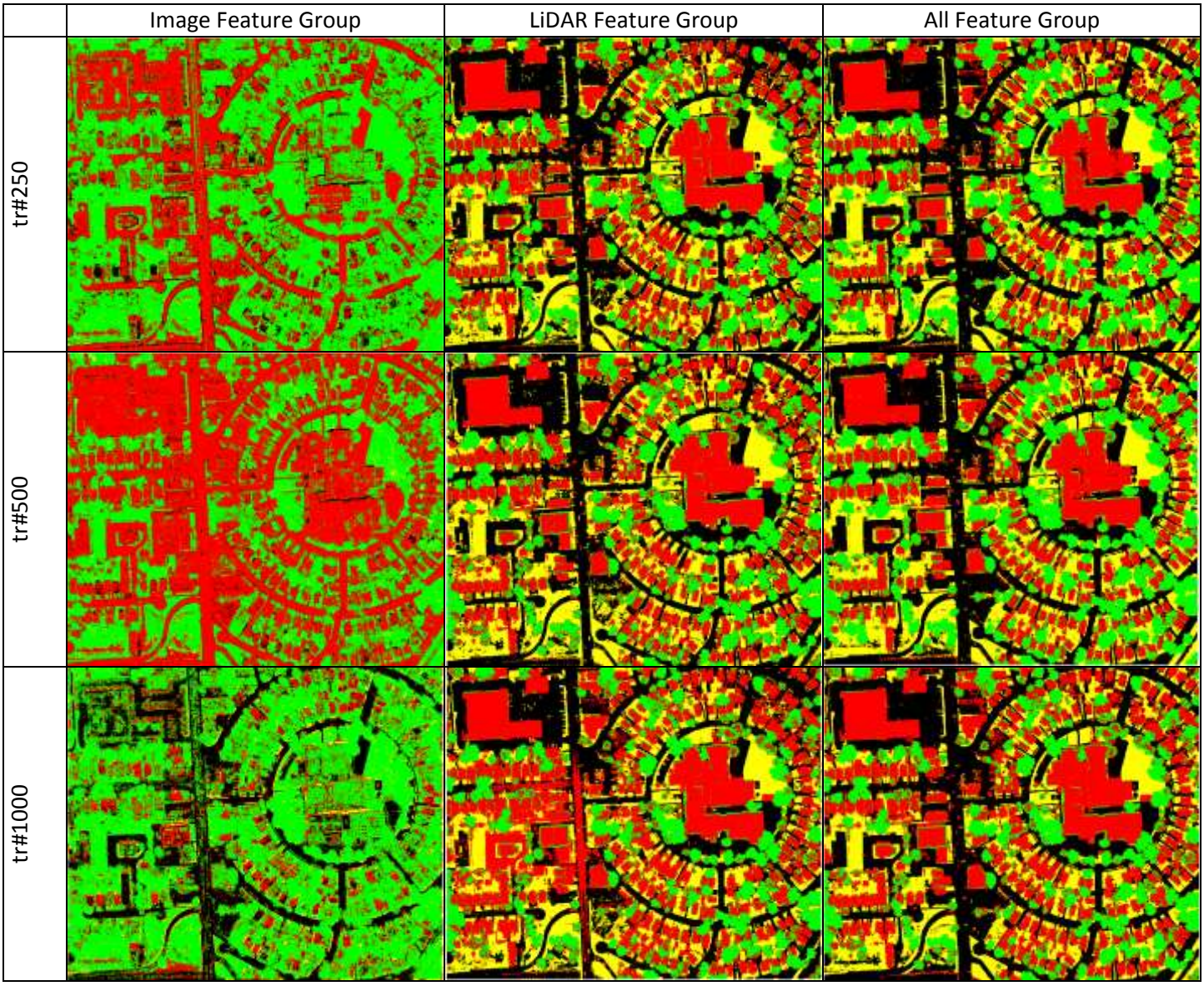
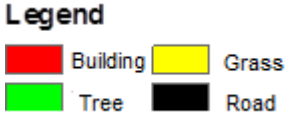


Figure 5.5 Classification maps achieved by RF-MLC



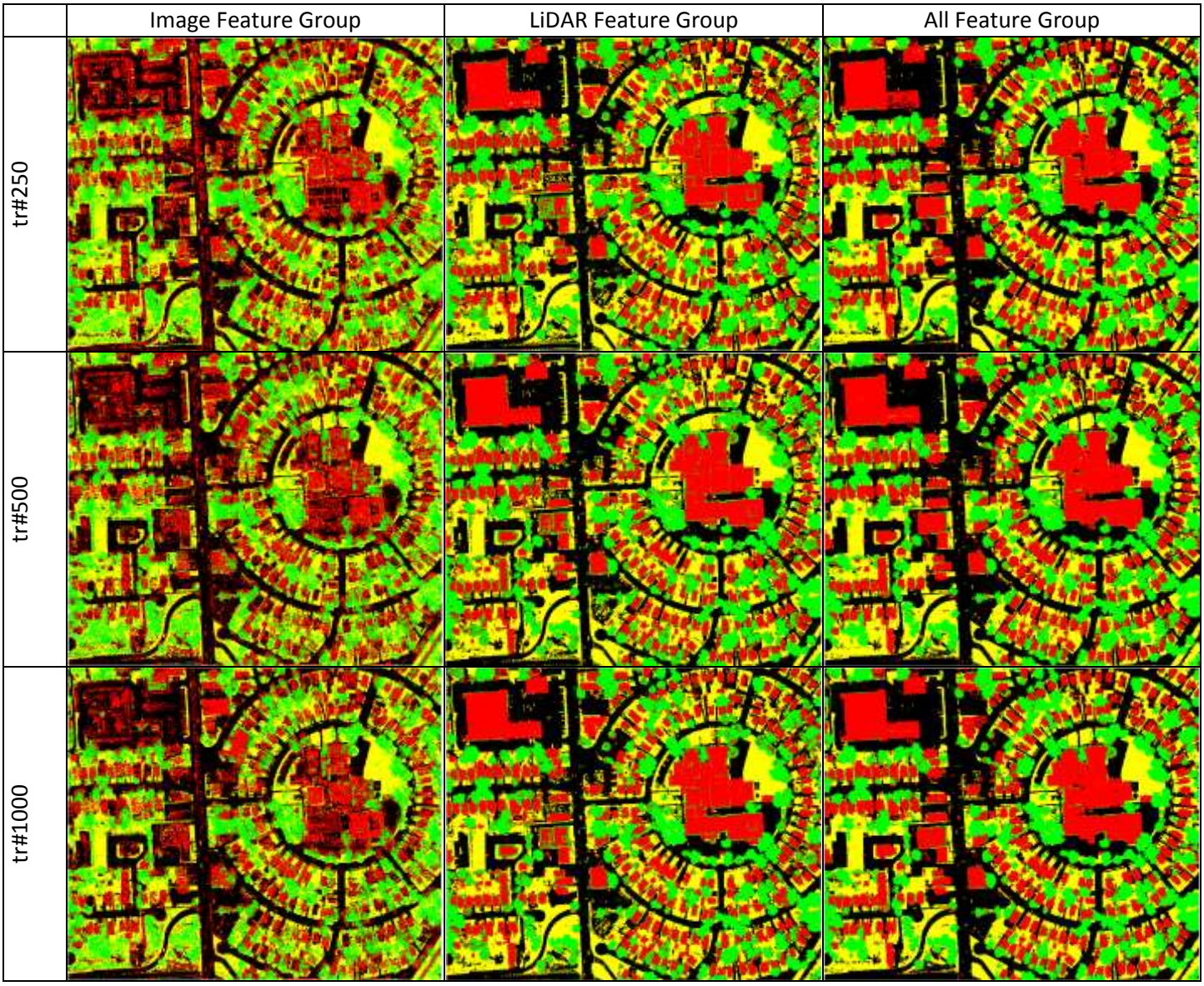
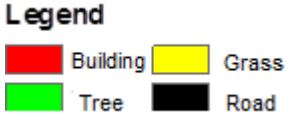


Figure 5.6 Classification maps achieved by PCA-SVM



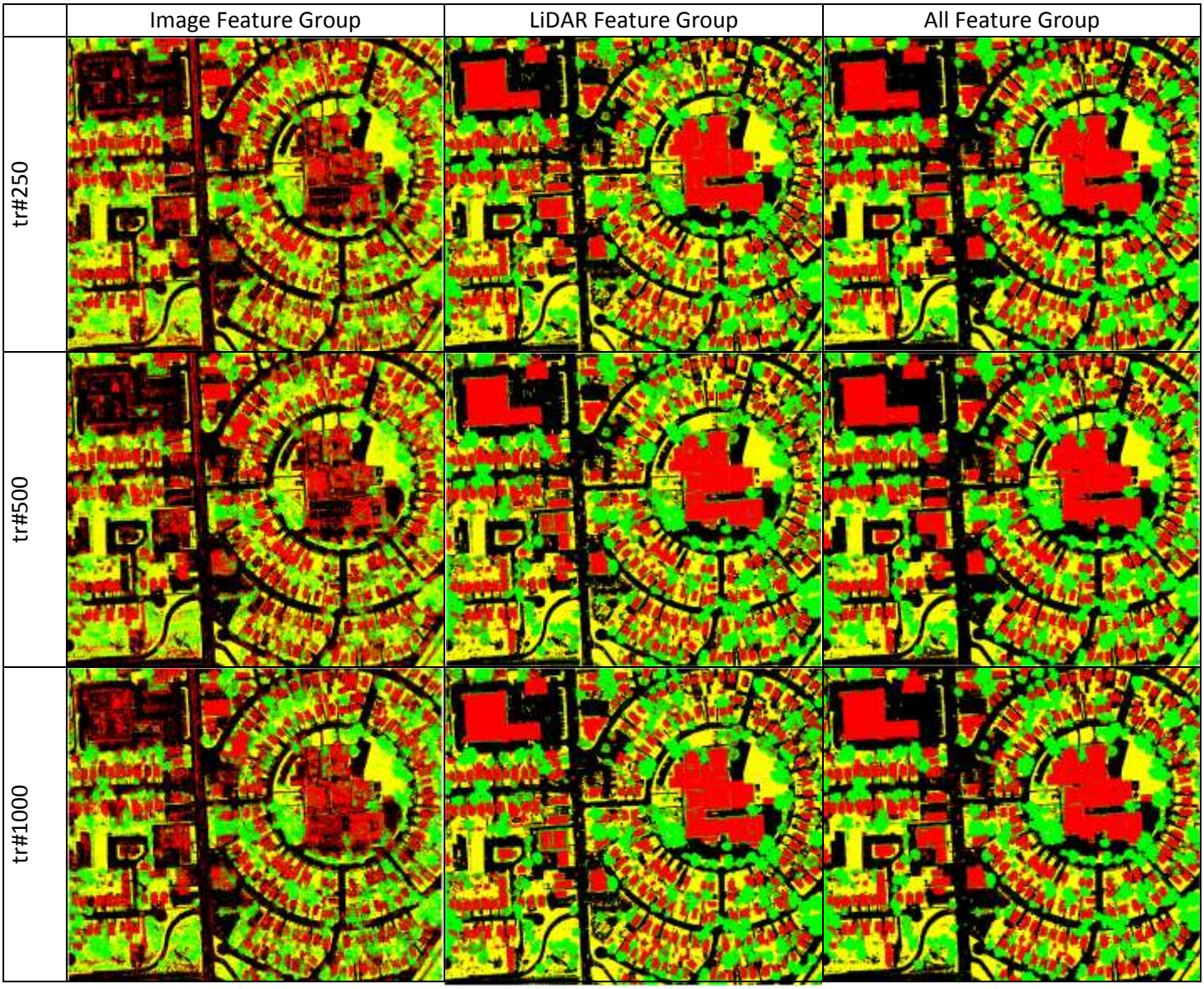
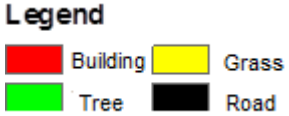


Figure 5.7 Classification maps achieved by RF-SVM



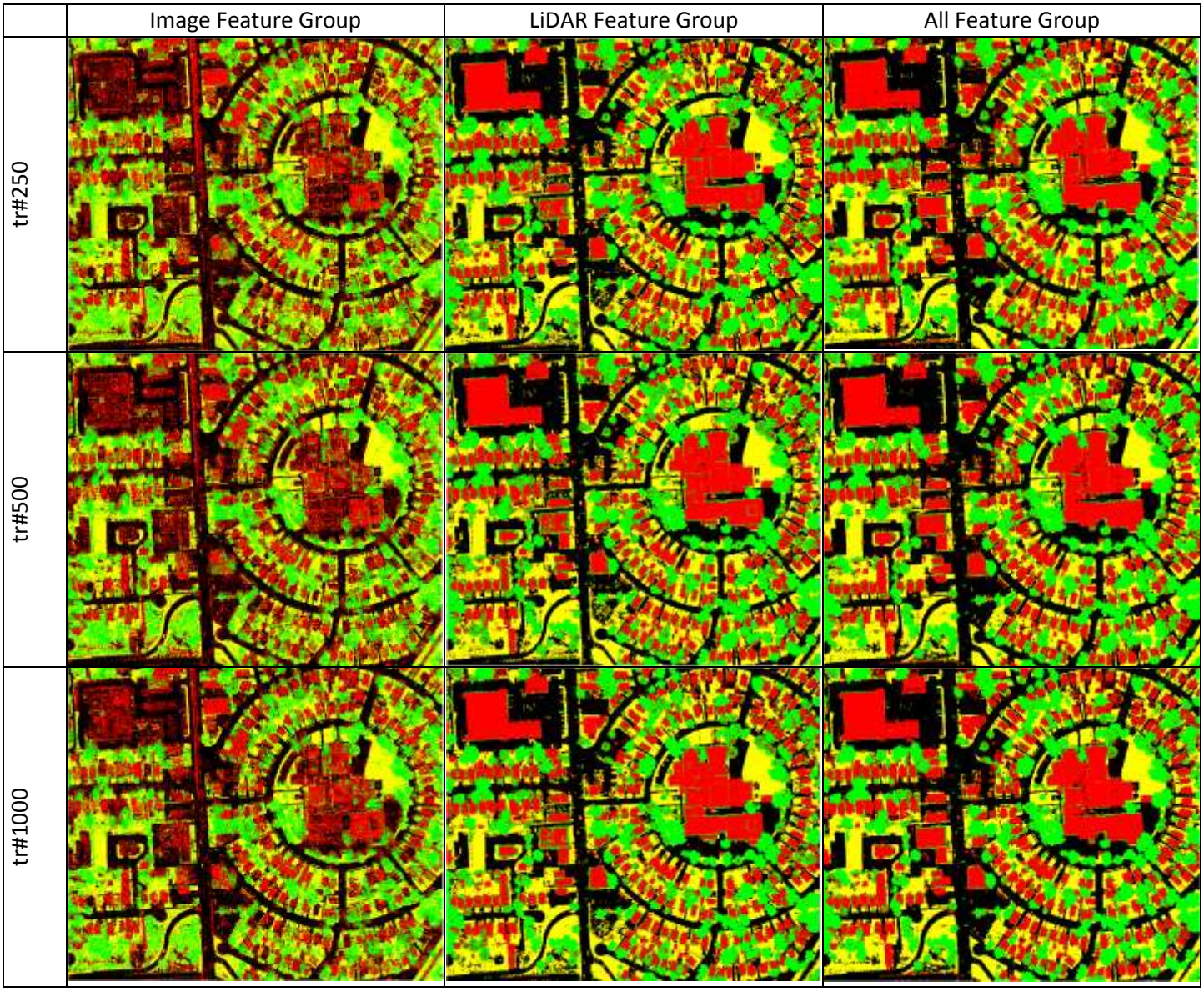
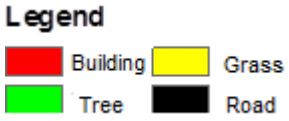


Figure 5.8 Classification maps achieved by PCA-RF



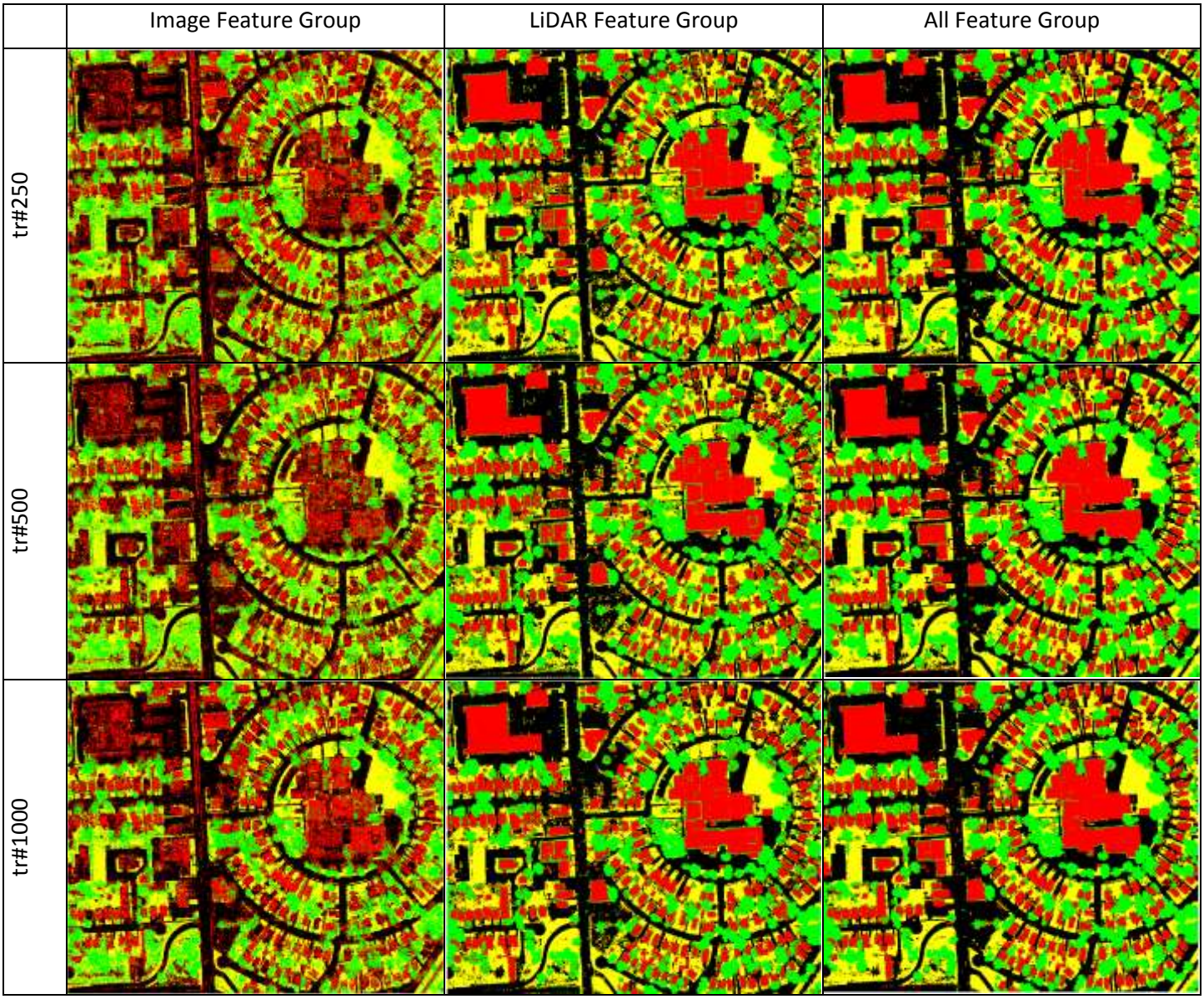


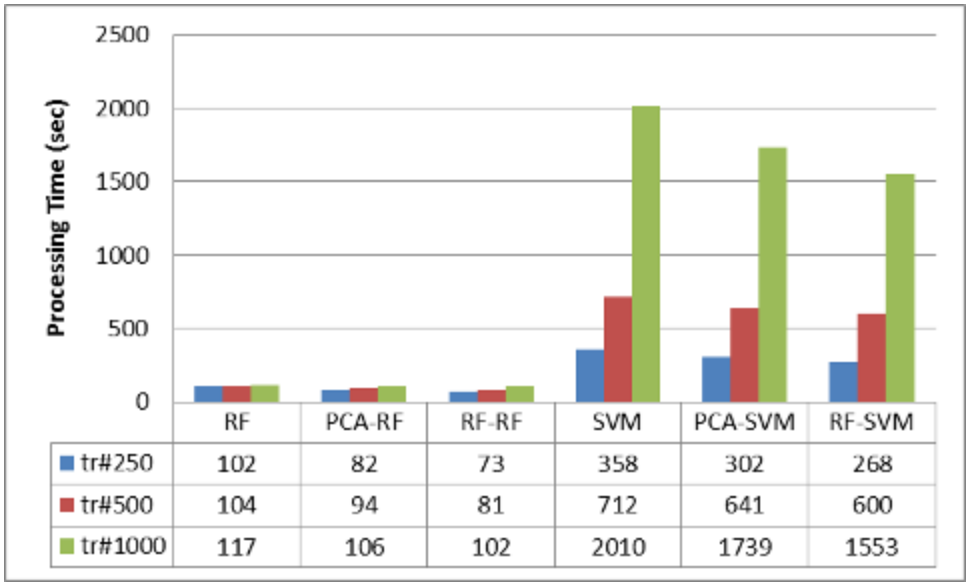
Figure 5.9 Classification maps achieved by RF-RF

5.4 Computation Time

Beside classification accuracy, computing speed is also an important criterion for evaluating the performance of classification methods. It took a few seconds for training the MLC in all training size levels. In this section, only the computation time of SVM based and RF based classifications, including SVM, PCA-SVM, RF-SVM, RF, PCA-RF, and RF-RF, is presented and compared in Tables 5-12 to 5-14. Table 5-12 presents the total processing time of classifications based on image feature group. Table 5-13 shows the computation time of LiDAR feature group classifications. Table 5-14 compares the total computation time of all feature group classifications.

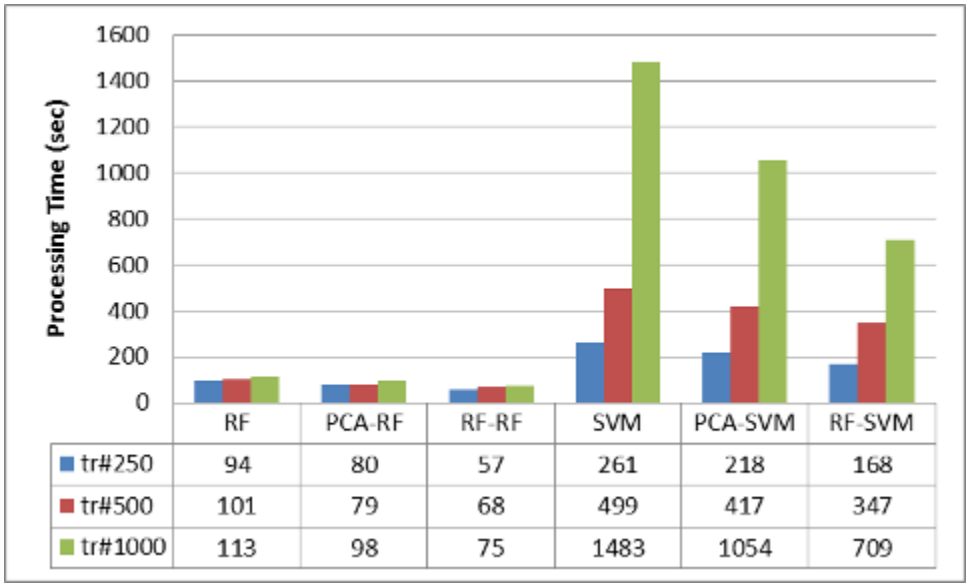
Regarding the computing speed of SVMs, the total results for parameter selection and classification processing time are reported. When the training size was doubled, the increases of SVM processing time grew exponentially. When the training size was increased from tr#250 to tr#1000, the classification time was much longer. Of SVM, PCA-SVM, and RF-SVM, RF-SVM took shortest processing time while SVM took longest. In contrast to SVM processing time, the RF classifications were much quicker. Because instead to transfer feature space to a higher dimensional one, the RF classifier try to construct many simple decision boundaries which are parallel to the feature axis from several decision trees. On all training data sizes, feature reduction by RF and PCA reduced the processing time for the classification. Nevertheless, the reduced degrees are lower on PCA than that on RF.

Table 5-12 Total processing time based on image feature group



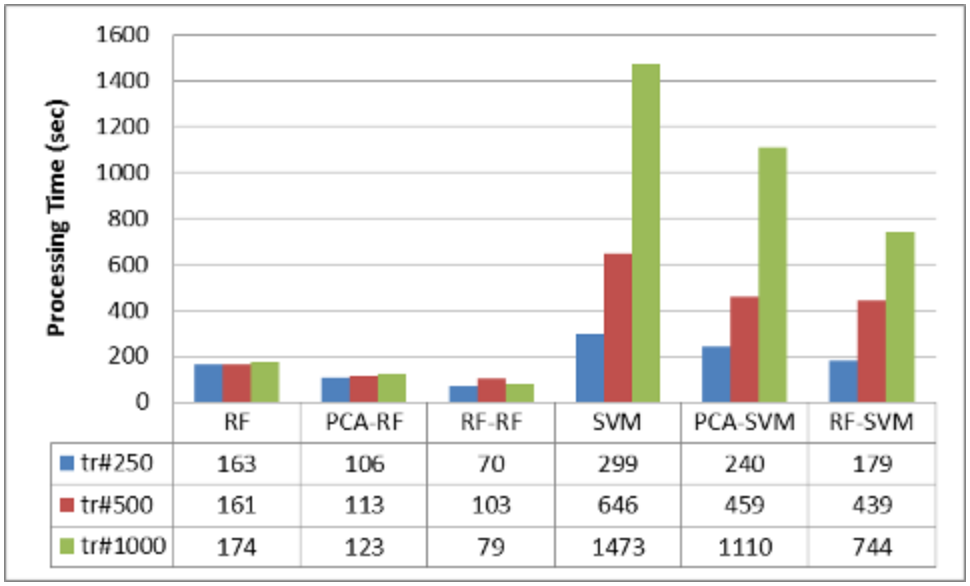
	RF	PCA-RF	RF-RF	SVM	PCA-SVM	RF-SVM
tr#250	102	82	73	358	302	268
tr#500	104	94	81	712	641	600
tr#1000	117	106	102	2010	1739	1553

Table 5-13 Total processing time based on LiDAR feature group



	RF	PCA-RF	RF-RF	SVM	PCA-SVM	RF-SVM
tr#250	94	80	57	261	218	168
tr#500	101	79	68	499	417	347
tr#1000	113	98	75	1483	1054	709

Table 5-14 Total processing time based on all feature group



	RF	PCA-RF	RF-RF	SVM	PCA-SVM	RF-SVM
tr#250	163	106	70	299	240	179
tr#500	161	113	103	646	459	439
tr#1000	174	123	79	1473	1110	744

5.5 Chapter Summary

In this chapter, all the urban land cover classification schemes presented in Section 4.5 have been investigated and compared in terms of accuracies, significance of difference, visible assessment, and computation time. Of all classification schemes, the MLC based classifications (MLC, PCA-MLC, and RF-MLC) had lower accuracies than the other approaches. These observations were in general agreement with the qualitative visual assessment. SVM based and RF based classifiers showed similar accuracies. However, RF based classifiers were much quicker than SVM based classifiers. In those applications which need a simple and quick algorithm for urban land cover classification, RF or RF-RF would be the best choice. Considering that the feature space is much smaller than the original one, we can overall assert that RF based feature selection is quite robust against the feature space dimensionality. When using a smaller number of features, the classifier structures can be simplified while presenting an equal classification performance. The assessment of the Producer's and User's accuracies underlined the general good performance of the proposed RF based feature selection and classification methods. Furthermore, results from RF based classifications showed a more positive balance between Producer's and User's accuracies of each class.

Chapter 6
Conclusions and Recommendations

This chapter presents the conclusion of this study, and the recommendations for future work. Section 6.1 gives the conclusion and implications of this thesis. Section 6.2 provides several suggestions for further research.

6.1 Conclusions

This study aims at investigating the performance of RF based feature selection and classification for airborne LiDAR and aerial image in urban area. Three different training sample sizes were collected to assess the effects of feature selection and classification methods. On the one hand, RF was considered as a feature selection method and compared with PCA based feature selection method based on the classification accuracy of three different methods (MLC, SVM, and RF). On the other hand, RF was considered as a classification method, and compared with other classification methods (MLC and SVM). The results showed that RF based feature selection and classification methods achieved promising results in terms of classification accuracy, ease of use, and computation time. Furthermore, the results showed that the integration of LiDAR data and aerial image can be applied to characterize complex urban scenes and derive much more accurate results than single image or single LiDAR data.

In this study, RF was found to be superior to PCA in feature selection for urban land cover classification. The main reason is that PCA relies on the whole data statistics, unlike RF which only based on the training samples. PCA minimizes the mean square error for a given number of features; however it is more desirable to generate reduced number of features which are focused on discrimination between classes. By projecting the original data into a lower dimension, PCA is suitable to discriminate the classes which are having the largest inter-variance (Cheriyadat and Bruce, 2003). When the classes contain a small difference in mean value, however, the PCA-derived data with reduced dimensionality may not be proper for discriminating them and thus degrading overall classification accuracy. The superiority of RF to PCA is that it relies on feature importance scores which are derived from decision tree classifiers. Thus, RF has the ability to select features focusing on discrimination between classes.

Of all three classification methods, SVM and RF based methods produced higher accuracies than MLC based methods. These observations generally agree with previous research where SVM and RF based classifications were found to be more accurate than conventional method like MLC, especially on small training size (Mantero et al., 2005; Ham et al., 2005). This is expected because, as discussed in Chapter 3, both SVM and RF are benefited from their advanced internal principles. Unlike MLC based on prior assumption made on data distribution, SVM adopt the structural risk minimization method to achieve minimum classification error on unseen data without a prior distribution assumption. RF involves bootstrapped sampling of the original training data for each decision tree, which could increase data diversity and reduce impact of outliers

due to reuse of training data. Moreover, SVM based and RF based classifications produced more balanced Producer's and User's accuracies than MLC based ones. When increasing the training data size, the accuracies of classifiers were increased in most cases. However, SVM based and RF based classifications performed more stably over different training sizes than MLC based classification.

Although SVM and RF based classifiers showed similar performance in terms of accuracy, the computation time of RF based classification is much shorter than SVM based classification. When the training size was doubled, the increases of SVM based computation time grew exponentially. In contrast, RF based classifiers took almost equal period over different training sample size. It implies that RF based classifiers are more effective and suitable to deal with large datasets than SVM.

In summary, this research described an encouraging finding for using LiDAR data and aerial imagery to map diverse urban land covers located in the City of Niagara Falls based on RF methods. The research has evaluated the different feature selection and classification methods. Based on the results, RF based feature selection is suited for reducing the data dimensionality of complex urban land cover types in the study area meanwhile reserving discrimination of different classes. In addition, RF based classification is accurate, effective, and user-friendly for urban land cover classification, even with small training size.

6.2 Recommendations for Future Work

This study also unveils some topics worth of further research efforts. The recommended future works are provided below to further improve the performance of RF based feature selection and classification methods.

First, although SVM and RF classifiers showed no significant difference from McNemar test, it is still worth trying to build a multiple system classifier by joining them together due to their varied basic assumptions (Waske et al., 2009b). Taking advantage of SVM and RF classifiers, the MSC has the potential to produce more robust classification results.

Second, all the classification results derived in this study used supervised pixel-based classifiers. In the future research, OBC could be tested and compared with the classification results in this study to further assess the performance of RF based methods.

Third, this study was applied to a relative flat urban area with no filtering of LiDAR data. However, it may be not appropriate for rough ground such as hilled areas. In the future research, digital terrain model (DTM) could be first generated from LiDAR data, and then used to produce nDSM (normalized DSM) by extracting DTM from DSM. In addition, multi-level classification system could be further built to classify different land cover type from different height level of nDSM.

Fourth, besides providing general feature importance ranking for all land cover types, RF method is able to produce importance measurements on individual land cover

type. Therefore, it has the potential to perform feature selection for particular interested land cover type.

Moreover, the proposed RF based method in this study mainly relies on the training data due to its nature of supervised algorithm. Thus, larger number of more accurate training samples could increase the performance of RF approaches. Guan et al. (2012) proposed a partially supervised method for hierarchal classification of urban land cover classification from LiDAR data and aerial imagery. In this study, a LiDAR-driven labeled image was used to generate larger training sample size with high-precision. In the future research, training data derived by partially supervised method will promote the capability of RF based feature selection and classification.

References

Agresti, A. *An Introduction to Categorical Data Analysis.* New York, NY: Wiley, 1996.

Alexander, C., K. Tansey, J. Kaduk, D. Holland, and N. J. Tate. "An approach to classification of airborne laser scanning point cloud data in an urban environment." *International Journal of Remote Sensing* 32, no. 24 (2011): 9151-9169.

Alpaydin, E.. "Support Vector Machines." In *Introduction to Machine Learning*, 218-226. Cambridge, MA: MIT Press, 2004.

Bartels, M., and H. Wei. "Maximum likelihood classification of LiDAR data incorporating multiple co-registered bands." *4th International Workshop on Pattern Recognition in Remote Sensing in Conjunction with the 18th International Conference on Pattern Recognition*, 2006: 17-20.

Bartels, M., and H. Wei. "Threshold-free object and ground point separation in LIDAR data." *Pattern Recognition Letters* 31 (2010): 1089-1099.

Batty, M. "The size, scale, and shape of cities." *Science* 319 (2008): 769-771.

Benediktsson, J. A., and J. R. Sveinsson. "Random forest classification of multisource remote sensing and geographic data." *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International* 2 (2004): 1049-1052.

Benz, U. C., P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen. "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS ready information." *ISPRS Journal of Photogrammetry & Remote Sensing* 58 (2004): 239-258.

Bhatta, B. *Analysis of Urban Growth and Sprawl from Remote Sensing Data.* Berlin, Heidelberg: Springer-Verlag, 2010.

Bilgin, G., S. Erturk, and T. Yildirim. "Segmentation of hyperspectral images via subtractive clustering and cluster validation using one-class support vector

machines." *IEEE Transactions on Geoscience and Remote Sensing* 49, no. 8 (2011): 2936-2944.

Bivand, R. S., E. J. Pebesma, and V. Gomez-Rubio. *Applied spatial data analysis with R.* NY: Springer, http://www.asdar-book.org/, 2008.

Blaschke, T. "Object based image analysis for remote sensing." *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (2010): 2-16.

Boser, B. E., I. M. Guyon, and V. N. Vapnik. "A training algorithm for optimal margin classifiers." *COLT '92: Proceeding of the Fifth Annual Workshop on Computational Learning Theory,New York, NY, USA:ACM Press*, 1992: 144-152.

Bottou, L., C. Cortes, J. Denker, H. Drucker, I. Guyon, L. Jackel, Y. LeCun, U. Muller, , E. Sackinger, P. Simard, and V. Vapnik. "Comparison of classifier methods: A case study in handwriting digit recognition." *Proc. Int. Conf. Pattern Recognition*, 1994: 77-87.

Bradley, J. V. *Distribution-Free Statistical Test.* Englewood Cliffs, New Jersey: Prentice-Hall, 1968.

Breiman, L. "Random forests." *Machine Learning* 45, no. 1 (2001): 5-32.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees.* Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1984.

Breiman, L.. "Manual for Setting Up, Using, and Understanding Random Forest V4.0." 2003. http://oz.berkeley.edu/users/breiman/Using_random_forests_v4.0.pdf (accessed 10 19, 2011).

Bruzzone, L., M. Marconcini, U. Wegmuller, and A. Wiesmann. "An advanced system for the automatic classification of multitemporal SAR images." *IEEE Transactions on Geoscience and Remote Sensing* 42, no. 6 (2004): 1321-1334.

Burges, C. J. C. "A tutorial on support vector machines for pattern recognition." *Data Mining and Knowledge Discovery* 2, no. 2 (1998): 121–167.

Camps-Valls , G., and L. Bruzzone. "Kernel-based methods for hyperspectral image classification." *IEEE Transactions on Geoscience and Remote Sensing* 43, no. 6 (2005): 1351-1362.

Chaelle, O., and S. Keerthi. "Multi-class feature selection with support vector machines." *In: Proceedings of the American Statistical Association, ASA, Denver,CO, USA, 3–7 August (on CD-ROM),* 2008.

Chan, J. C.-W., and D. Paelinckx. "Evaluation of random forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery." *Remote Sensing of Environment* 112 (2008): 2999-3011.

Chan, J. C.-W., N. Laporte, and R. S. DeFries. "Texture classification of logged forests in tropical Africa using machine-learning algorithms." *International Journal of Remote Sensing* 24, no. 6 (2003): 1401-1407.

Chang, C.-C., and C.-J. Lin. "LIBSVM: A Library for Support Vector Machines." *ACM Transactions on Intelligent Systems and Technology* 2, no. 3 (2011): 1-27.

Chehata, N., L. Guo, and C. Mallet. "Airborne Lidar feature selection for urban classification using random forests." In *Laser Scanning 2009, IAPRS, Vol.XXXVIII, Part 3/W8*, edited by F. Bretar, M. Pierror-Deseilligny and G. Vosselman. 2009.

Chen, D., D. A. Stow, and P. Gong. "Examining the effect of spatial resolution and texture window size on classification accuracy: An urban environment case." *International Journal of Remote Sensing* 25 (2004): 2177-2192.

Chen, L. C., T.A. Teo, C. Y. Kuo, and J. Y. Rau. "Shaping polyhedral buildings by the fusion of vector maps and lidar point clouds." *Photogrammeric Engineering & Remote Sensing* 79, no. 9 (2008): 1147-1157.

Chen, Y. W., and C. J. Lin . "Combining SVMs with various feature selection strategies." In *Feature Extraction, Foundations and Applications*, edited by I. Guyon, S. Gunn, M. Nikravesh and L.A. Zadeh. Physica-Verlag, USA: Springer, 2006.

Cheriyadat, A., and L. M. Bruce . "Why principal component analysis is not an appropriate feature extraction method for hyperspectral data." *Proceedings of the Geoscience and Remote Sensing Symposium 2003 (IGARSS '03),*, 2003: 3420-3422.

Clausi, D. A. "An analysis of co-occurrence texture statistics as a function of grey-level quantization." *Canadian Journal of Remote Sensing* 28 (2002): 45-62.

Clode, S., F. Rottensteinerb, P. Kootsookosc, and E. Zelniker. "Detection and vectorization of roads from Lidar data." *Photogrammeric Engineering and Remote Sensing* 73, no. 5 (2007): 517-535.

Congalton, R. G. *The Use of Discrete Multivariate Analysis for the Assessment of Landsat Classification Accuracy.* Blacksburg, VA: Virginia Polytechnic Institute and Sate University, Mater's thesis, 1981.

Congalton, R. G., and K. Green. *Assessing the Accuracy of Remotely Sensed Data : Principles and Practices*. Boca Raton: CRC Press, 2009.

Cortes, C., and V. Vapnik. "Support-vector network." *Machine Learning* 20, no. 3 (1995): 273-297.

Cristianini, N., and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* Cambridge, UK: Cambridge University Press, 2000.

Dalponte, M., L. Bruzzone, and D. Gianelle. "Fusion of hyperspectral and LiDAR remote sensing data for classification of complex forest areas." *IEEE Transactions on Geoscience and Remote Sensing* 46, no. 5 (2008): 1416-1427.

Deering, D. W., J. W. Rouse, R. H. Haas, and J. A. Schell. "Measuring forage production of grazing units from Landsat MSS data." *Proceeding of 10th International Symposium on Remote Sensing of Environment ERIM 2*, 1975: 1169-1178.

Definiens. *eCognition Developer 8: User Guide*. München, Germany: Definiens AG, 2009.

Defries, R. S., and J. C. Chan. "Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data." *Remote Sensing of Environment* 74 (2000): 503-515.

Del Frate, F., F. Pacifici, G. Schiavon, and C. Solimini. "Use of Neural Networks for automatic classification from high-resolution images." *IEEE Transactions on Geoscience and Remote Sensing* 45, no. 4 (2007): 800-809.

Dixon, B., and N. Candade. "Multispectral landuse classification using neural networks and support vector machines: One or the other, or both?" *International Journal of Remote Sensing* 29, no. 4 (2008): 1185-1206.

Dorninger, P., and N. Pfeifer. "A Comperhensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds." *Sensors* 8 (2008): 7323-7343.

Duda, R. O., P. E. Hart, and D. G Stork. *Pattern Classification.* John Wiley & Sons, Inc., 2001.

Duda, T., and M. J. Canty. "Unsupervised classification of satellite imagery: Choosing a good algorithm." *International Journal of Remote Sensing* 23, no. 11 (2002): 2193-2212.

Estep, L., G. Terrie, and B. Davis. "Crop stress detection using AVIRIS hyperspectral imagery and artificial neural networks." *International Journal of Remote Sensing* 25 (2004): 4999-5004.

Filippi, A. M., and J. R. Jensen. "Fuzzy learning vector quantization for hyperspectral coastal vegetation classification." *Remote Sensing of Environment* 100 (2006): 512-530.

Fletcher, T.. "Support Vector Mahines Explained." 2009. http://www.tristanfletcher.co.uk/SVM%20Explained.pdf (accessed 10 19, 2011).

Foody, G. M. "On training and evaluation of SVM for remote sensing applications." In *Kernel Methods for Remote Sensing Data Analysis*, edited by Gustavo Camps-Valls and L. Bruzzone, 85-109. UK: Jonh Wiley&Sons Ltd, 2009.

Foody, G. M. "Supervised image classification by MLP and RBF neural networks with and without an exhaustively defined set of classes." *International Journal of Remote Sensing* 25 (2004a): 3091-3104.

Foody, G. M. "Thematic map comparison : Evaluating the statistical significance of differences in classification accuracy." *Photogrammetric Engineering & Remote Sensing* 70, no. 5 (2004b): 627-633.

Foody, G. M., A. Mathur, C. Sanchez-Hernandez, and D. S. Boyd. "Training set size requirements for the classification of a specific class." *Remote Sensing of Environment*, no. 104 (2006): 1-14.

Foody, G. M., and A. Mathur. "The use of small training sets containing mixed pixels for accurate hard image classification training on mixed spectral responses for classification by a SVM." *Remote Sensing of Environment*, no. 103 (2006): 179-189.

Foody, G.M., and A. Mathur. "A relative evaluation of multiclass image classification by support vector machines." *IEEE Transactions on Geoscience and Remote Sensing* 42 (2004): 1335-1343.

Friedl, M. A., and C. E. Brodley. "Decision tree classification of land cover from remotely sensed data." *Remote Sensing of Environment* 61 (1997): 399-409.

Friedman, J. H. "Another Approach to Polychotomous Classification." 1996. http://www-stat.stanford.edu/~jhf/ftp/poly.pdf (accessed 10 19, 2011).

Frohlich, H., O. Chapelle, and B. Scholkopf. "Feature selection for support vector machines using genetic algorithms." *International Journal on Artificial Intelligence Tools* 13, no. 4 (2004): 791-800.

Fuller, D. O. "Remote detection of invasive Melaleuca trees (Melaleuca quinquenervia) in South Florida with multispectral IKONOS imagery." *International Journal of Remote Sensing* 26, no. 5 (2005): 1057-1063.

Garcá-Gutiérrez, J., F. Martínez-Álvarez, and J. C. Riquelme. "Using remote data mining on LIDAR and imagery fusion data to develop land cover maps." *Lecture Notes in Computer Science* 6096 (2010): 378-387.

García-Gutiérrez, J., L. Gonçalves-Seco, and J.C. Riquelme-Santos. "Automatic environmental quality assessment for mixed-land zones using lidar and intelligent techniques." *Expert Systems with Applications: An International Journal* 38, no. 6 (2011): 6805-6813.

Geneletti, D., and B. G. H. Gorte. "A method for object-oriented land cover classification combining Landsat TM data and aerial photographs." *International Journal of Remote Sensing* 24 (2003): 1273-1286.

Germaine, K. A., and M. C. Hung. "Delineation of impervious surface from multispectral imagery and lidar incorporating knowledge based expert system rules." *Photogrammetric Engineering & Remote Sensing* 77, no. 1 (2011): 75-85.

Ghimire, B., J. Rogan, and J. Miller. "Contextual land-cover classification: Incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic." *Remote Sensing Letters* 1 (2010): 45-54.

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson. "Random forests for land cover classification." *Pattern Recognition Letters* 27, no. 4 (2006): 294-300.

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson. "Random forest classification of multisource remote sensing and geographic data." *In: IEEEInternational Geoscience and Remote Sensing Symposium, Anchorage, AK, 20-24 September*, 2004: 1049-1052.

Gitas, I. Z., G. H. Mitri, and G. Ventura. "Object-based image classification for burned area mapping of Creus Cape Spain, using NOAA-AVHRR imagery." *Remote Sensing of Environment* 92 (2004): 409-413.

Gómez-Chova, L., J. Muñoz-Marí, V. Laparra, J. Malo-López, and G. Camps-Valls. "A review of kernel methods in remote sensing data analysis." In *Optical Remote Sensing, Augmented Vision and Reality Series: 3*, edited by S. Prasad, L. M. Bruce and J. Chanussot, 171-206. Verlag Berlin Heidelberg: Springer, 2011.

Goodwin, N. R., N. C. Coops, T. R. Tooke, A. Christen, and J. A. Voogt. "Characterizing urban surface cover and structure with airborne lidar technology." *Canadian Journey of Remote Sensing* 35, no. 3 (2009): 297-309.

Grimm, N. B., S. H. Faeth, N. E. Golubiewski, C. L. Redman, J. Wu, X. Bai, and J. M. Briggs. "Global change and ecology of cites." *Science* 319 (2008): 756-760.

Gross, H., and U. Thoennessen. "Extraction of lines from laser point clouds." *ISPRS Conference Photogrammetric Image Analysis (PIA)* 36, no. 3A (2006): 87-91.

Guan, H., Z. Ji, L. Zhong, J. Li, and Q. Ren. "Partially supervised hierarchal classification for urban features from Lidar data with aerial imagery." *International Journal of Remote Sensing* Accepted (2012).

Guo, Li, N. Chehata, C. Mallet, and S. Boukir. "Relevance of airborne Lidar and multispectral image data for urban scene classification using Random Forests." *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (2011): 56-66.

Guyon, I., S. Gunn, M. Nikravesh, and L. Zadeh. *Feature Extraction. Foundations and Applications.* Physica-Verlag, USA: Springer, 2006.

Hall, D., and G. Ball. *Isodata: A Novel Method of Data Analysis and Pattern Classification (Technical Report).* Stanford Research Institute, 1965.

Ham, J., Y. Chen, and M. M. Crawford. "Investigation of the random forest framework for classification of hyperspectral data." *IEEE Transactions on Geoscience and Remote Sensing* 43, no. 3 (2005): 492-501.

Han, J,, and M. Kamber. *Data Mining: Concepts and Techniques.* San Francisco, CA: Morgan Kaufmann Publishers, 2001.

Hansen, M., R. Dubayah, and R. Defries. "Classification trees: An alternative to traditional land cover classifiers." *International Journal of Remote Sensing* 17 (1996): 1075-1081.

Hao, Z., Y. Zhang, J. Liu, and S. Ji . "Automatic building detection using airborne LIDAR data." *International Forum on Information Technology and Applications* 3 (2009): 668-671.

Haralick, R. M. "Statistical and structural approaches to texture." *Proceedings of the IEEE* 67 (1979): 786-804.

Haykin, S. *Neural Networks: A Comprehensive Foundations.* 2nd. Upper Saddle River, New Jersey: Prentice Hall, 1999.

Hepinstall, J.A., M. Alberti, and J.M. Marzluff. "Predicting land cover change and avian community reponses in rapidly urbanizing environments." *Landscape Ecology* 23 (2008): 1257-1276.

Herold, M., X. Liu, and K. C. Clarke. "Spatial metrics and image texture for mapping urban land use." *Photogrammetric Engineering and Remote Sensing* 69 (2003): 991-1001.

Hodgson, M. E., J. R. Jensen, J. A. Tullis, K.D. Riordan, and C.M. Archer. "Synergistic use lidar and color aerial photography for mapping urban parcel imperviousness." *Photogrammetric Engineering and Remote Sensing* 69 (2003): 973-980.

Hofle, B., and N. Pfeifer. “Correction of laser scanning intensity data: Data and model-driven approaches.” *ISPRS Journal of Photogrammetry & Remote Sensing* 62 (2007): 415-433.

Hofle, B., M. Hollaus, and J. Hagenauer. "Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data." *ISPRS Journal of Photogrammetry and Remote Sensing* 67 (2012): 134-147.

Horning, N. "Random Forests : An algorithm for image classification and generation of continuous fields data sets." *Proceeding of International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences*, 2010.

Hsu, C.-W., C.-C. Chang, and C.-J. Lin. "A Practical Guide to Support Vector Classification." 2010. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf (accessed 10 19, 2011).

Huang, C., L. S. Davis, and J. R. G. Townshend. "An assessment of support vector machines for land cover classification." *Internatioanl Journal of Remote Sensing* 23, no. 4 (2002): 725-749.

Huang, C.-L., and C.-J. Wang. "A GA-based feature selection and parameters optimization for support vector machines." *Expert Systems with Application* 31 (2006): 231-240.

Huang, M.-J., S.-W. Shyue, L.-H. Lee, and C.-C. Kao. "A knowledge-based approach to urban feature classification using aerial imagery with Lidar data." *Photogrammetric Engineering & Remote Sensing* 74, no. 12 (2008): 1473-1485.

Huang, X., L. Zhang, and W. Gong. "Information fusion of aerial images and Lidar data in urban areas: Vetor-stacking, re-classification and post-processing approaches." *International Journal of Remote Sensing* 32, no. 1 (2011): 69-84.

Huber, M., W. Schickler, S. Hinz, and A. Baumgartner. "Fusion of Lidar data and aerial imagery for automatic reconstruction of building surfaces ." *2nd Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, 2003.

Hughes, G. F. "On the mean accuracy of statistical pattern recognition." *IEEE Trans. Inform. Theory* IT, no. 14 (1968): 55-63.

Hung, M., and M. K. Ridd. "A subpixel classifier for urban land-cover mapping based on a maximum-likelihood approach and expert system rules." *Photogrammetric Engineering and Remote Sensing* 68 (2002): 1173-1180.

Hussain, E., S. Ural, K. Kim, C.-S. Fu, and J. Shan. "Building extraction and rubble mapping for City Port-au-Prince Post-2010 earthquake with GeoEye-1 imagery and Lidar data." *Photogrammetric Engineering & Remote Sensing* 77, no. 10 (2011): 1011-1023.

Ismail, R. *Remote Sensing of Forest Health: The Detection and Mapping of Pinus Patula Trees Infested by Sirex Noctilio.* PhD Thesis, University of KwaZulu-Natal, Pietermaritzburg, 2009.

Jain, A. K., R. P. W. Duin, and J. C. Mao. "Statistical pattern recognition: A review." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, no. 1 (2000): 4-37.

Janz, A., S. van der Linden, B. Waske, and P. Hoster. "imageSVM: A user-oriented tool for advanced classification of hyperspectral data using support vector machines." In *Proceeding of 5th EARSeL Workshop on Imaging Spectroscopy*, edited by I. Reusen and J. Cools. 2007.

Jensen, John R. "Thematic Map Accuracy Assessment." In *Introductory Digital Image Processing: A Remote Sensing Perspective*, 495-515. Upper Saddle River: Pearson Prentice Hall, 2005.

Jiang, H., J. R. Strittholt, P. A. Frost, and N. C. Slosser. "The classification of late seral forests in the Pacific Northwest USA using Landsat ETM + imagery." *Remote Sensing of Environment* 91 (2004): 320-331.

Joelsson, S. R., J. A. Benediktsson, and R. Sveinsson. "Random forest classification of remote sensing data." In *Image Processing For Remote Sensing*, edited by C.-H. Chen, 61-78. Boca Racon, FL: CRC Press, 2008.

Jones, T. G., N. C. Coops, and T. Sharma. "Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada." *Remote Sensing of Environment* 114, no. 12 (2010): 2841-2852.

Keerthi, S. S., and C.-J. Lin. "Asymptotic behaviors of support vector machines with Gaussian kernel." *Neural Computation* 15, no. 7 (2003): 1667-1689.

Keramitsoglou , I., H. Sarimve, C. T. Kiranoudisb , and N. Sifakisc. "Radial basis function neural networks classification using very high spatial resolution satellite imagery: An application to the habitat area of Lake Kerkini (Greece)." *International Journal of Remote Sensing* 26 (2005): 1861-1880.

Ketting, R. L., and D. A. Landgrebe. "Classification of multispectral image data by extraction and classification of homogenous objects." *IEEE Transactions on Geoscience and Remote Sensing* 14 (1976): 19-26.

Khoshelham, K., C. Nardinocchi, E. Frontoni, A. Mancini, and P. Zingaretti. "Performance evaluation of automated approaches to building detection in multi-source aerial data." *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (2010): 123-133.

Kim, S., T. Hinckley, and D. Briggs. "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 (2011): 3329-3342.

Knerr, S., L. Personnaz, and G. Dreyfus. "Single-layer learning revisited: A stepwise procedure for building and training a neural network." In *Neurocomputing: Algorithms, Architectures and Applications*, edited by J. Fogelman. Verlag: Springer, 1990.

Koetz, B., F. Morsdorf, S. van der Linden, T. Curt, and B. Allgöwer. "Multi-source land cover classification for forest fire management based on imaging spectrometry and lidar data." *Forest Ecology and Management* 256 (2008): 263-271.

Koltunov, A., and E. Bendor. "A new approach for spectral feature extraction and for unsupervised classification of hyperspectral data based on the Gaussian mixture model." *Remote Sensing Reviews* 20 (2001): 123-167.

Koltunov, A., and E. Bendor. "Mixture density separation as a tool for high-quality interpretation of multi-source remote sensing data and related issues." *International Journal of Remote Sensing* 25 (2004): 3275-3299.

Kontoes, C. C., V. Raptis, M. O. Lautner, and R. Oberstadler. "The potential of kernel classification techniques for land use mapping in urban areas using 5m-spatia resolution IRS-1C imagery." *International Journal of Remote Sensing* 21 (2000): 3145-3151.

Kontoes, C. C., and D. Rokos. "The integration of spatial context information in an experimental knowledge based system and the supervised relaxation algorithm: two successful approaches to improving SPOT-XS classification." *International Journal of Remote Sensing* 17 (1996): 3039-3106.

Kreßel, U. "Pairwise classification and support vector machines." In *Advances in Kernel Methods - Support Vector Learning*, edited by B. Schölkopf, C. J.C. Burges and A. J. Smola, 255–268. Cambridge, MA: MIT Press, 1999.

Kurnaz, M. N., Z. Dokur, and T. Olmez. "Segmentation of remote-sensing images by incremental neural network." *Pattern Recognition Letters* 26 (2005): 1096-1104.

Laliberte, A., D. Browning, and A. Rango. "A comparison of three feature selection methods for object-based classification of sub-decimeter resolution UltraCam-L

imagery." *International Journal of Applied Earth Observation and Geoinformation* 15 (2012): 70-78.

Landis, J., and G. Koch. "The measurement of observer agreement for categorical data." *Biometrics* 33 (1977): 159-174.

Lee, C., and D. A. Landgrebe. "Feature extraction based on decision boundaries." *IEEE Transaction on Pattern Analysis and Machine Intelligence* 15 (1993): 388-400.

Lee, D. H., K. M. Lee, and S. U. Lee. "Fusion of lidar and imagery for reliable building extraction." *Photogrammeric Engineering and Remote Sensing* 74, no. 2 (2008): 215-225.

Lewin-Koh, N. J., and R. Bivand. "Maptools: Tools for reading and handling spatial objects." 08 08, 2011. http://cran.r-project.org/web/packages/maptools/index.html (accessed 11 17, 2011).

Li, C. H., B. C. Kuo, C. T. Lin, and C. S. Huang. "A spatial–contextual support vector machine for remotely sensed image classification." *IEEE Transactions on Geoscience and Remote Sensing* 50, no. 3 (2012): 784-799.

Li, H., H. Gu, Y. Han, and J. Yang. "Fusion of high-resolution aerial imagery and lidar data for object-oriented urban land-cover classification based on SVM." *In Proceedings of the ISPRS Workshop on Updating Geospatial Databases with Imagery & The 5th ISPRS Workshop on DMGISs*, 2007: 179-183.

Li, H., Y. Wang, Y. Li, and X. Wang. "Pixel-unmixing moderate-resolution remote sensing imagery using pairwise coupling support vector machines: A case study." *IEEE Transactions on Geoscience and Remote Sensing* 49, no. 11 (2011): 4298-4307.

Li, H., L. Di, X. Huang, and D. Li. "Laser intensity used in classification of Lidar point cloud data." *IGRASS* 2 (2008): 1140-1143.

Liaw, A., and M. Wiener. "Classification and regression by randomForest." *R News* 2, no. 3 (2002): 18-22.

Lin, H.-T., and C.-J. Lin. "A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods, technical report, Department of Computer

Science, National Taiwan University." 2003. http://www.csie.ntu.edu.tw/~cjlin/papers/tanh.pdf (accessed 10 19, 2011).

Liu, H., and H. Motoda. *Computational Methods of Feature Selection.* Roca Raton, FL: Taylor & Francis Group, LLC, 2008.

Liu, H., and L. Yu. "Toward integrating feature selection algorithms for classification and clustering." *IEEE Transactions on Knowledge and Data Engineering* 17, no. 4 (2005): 491-502.

Liu, W., S. Gopal, and C. E. Woodcock. "Uncertainty and confidence in land cover classification using a hybrid classifier approach." *Photogrammeric Enginnering and Remote Sensing* 70 (2004): 963-971.

Lo, C. P. "The application of geospatial technology to urban morphological research." *Urban Morphology* 11 (2007): 81-90.

Lodha, S. K., E. J. Kreps, D. P. Helmbold, and D. Fitzpatirck. "Aerial lidar data classification using support vector machines (SVM)." *Proceedings of the Third International Symposium on 3D Data Processing, Visualization and Transmission,* 2006.

Longepe, N., P. Rakwatin, O. Isoquchi, M. Shimada, Y. Uryu, and K. yulianto. "Assessment of ALOS PALSAR 50 m orthorectified FBD data for regional land cover classification by support vector machines." *IEEE Transactions on Geoscience and Remote Sensing* 49, no. 6 (2011): 2135-2150.

Loosvelt, L., J. Peters, H. Skriver, B. De Baets, and N.E.C. Verhoest. "Impact of reducing Polarimetric SAR input on the uncertainty of crop classifications based on the random forests algorithm." *IEEE Transactions on Geoscience and Remote Sensing* Early Access Articles (2012): 1-16.

Lu, D. and Q. Weng. " A survey of image classification methods and techniques for improving classification performance." *Inhternational Journal of Remote Sensing* 28, no.5 (2007): 823-870.

Lu, S., K. OKI, Y. SHIMIZU, and K. OMASA. "Comparison between several feature extraction classification methods for mapping complicated agricultural land use

patches using airborne hyperspectral data." *International Journal of Remote Sensing* 28, no. 5 (2007): 963-984.

Luo, Y., T. Jiang, S. Gao, and X. Wang. "A new method of building footprints detection using airborne laser scanning data and multispectral image." *5th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Optoelectronic Material and Devices for Detector, Imager, Display, and Energy Conversion Technology*, 2010.

MacQueen, J. "Some methods for classification and analysis of multivariate observations." *Fifth Berkeley Symposium on Mathematics,Statistics and Probability*, 1967: 281–296.

Mallet, C., F. Bretar, and U. Soergel. "Analysis of full-waveform lidar data for classification of urban areas." *Photogrammetrie Fernerkundung Geoinformation* 5 (2008): 337-349.

Mallet, C., F. Bretar, M. Roux, U. Soergel, and C. Heipke. "Relevance assessment of full-waveform lidar data for urban area classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (2011): S71-S84.

Malpica, J. A., and M. C. Alonso. "Urban changes with satellite imagery and Lidar data." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science* XXXVIII, no. 8 (2010): 853-858.

Mantero, P., G. Moser, and S. B. Serpico. "Partially supervised classification of remote sensing images through SVM-based probability density estimation." *IEEE Transactions on Geoscience and Remote Sensing* 43, no. 3 (2005): 559-570.

Martinuzzi, S., L. A. Vierling, W. A. Gould, M. J. Falkowski, J. S. Evans, A. T. Hudak, and K. T. Vierling,. "Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability." *Remote Sensing of Environment* 113, no. 12 (2009): 2533-2546.

Mas, J. F., and J. J. Flores. "The application of artificial neural networks to the analysis of remotely sensed data." *International Journal of Remote Sensing* 29, no. 3 (2008): 617-663.

Melgani, F., and L. Bruzzone. "Classification of hyperspectral remote sensing images with support vector machines." *IEEE Transactions on Geoscience and Remote Sensing* 42, no. 8 (2004): 1778-1790.

Meng, X., L. Wang, J. L. Silivan-Cardenas, and N. Currit. "A multi-directional ground filtering algorithm for airborne Lidar." *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (2009): 117-124.

Meng, X., N. Currit, L. Wang, and X. Yang. "Detect residential buildings from Lidar and aerial photographs through object-oriented land-use classification." *Photogrammetric Engineering and Remote Sensing* 78, no. 1 (2012): 35-44.

Mercer, J. "Functions of positive and negative type and their connection with the theory of integral equations." *Philosophical Transactions of the Royal Society A* 209 (1909): 415-446.

Mitchell, T. M. *Machine Learning.* USA: WCB/McGraw-Hill, 1997.

Mitra, P., C. Murthy, and S. Pal. "Unsupervised feature selection using feature similarity." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, no. 3 (2002): 301-312.

Mittelbach, F.G., and M.I. Schneider. "Remote sensing: With special reference to urban and regional transportation." *Annals of Regional Science* 5 (2005): 61-72.

Mongus, D., and B. Zalik. "Parameter-free ground filtering of LiDAR data for automatic DTM generation." *ISPRS Journal of Photogrammetry and Remote Sensing* 67 (2012): 1-12.

Moustakidis, S., G. Mallinis, N. Koutsias, and J. B. Theocharis. "SVM-based fuzzy decision trees for classification of high spatial resolution remote sensing images." *IEEE Transactions on Geoscience and Remote Sensing* 50, no. 1 (2012): 149-169.

Muchoney, D., and A. Strahler. "Regional vegetation mapping and direct land surface parameterization from remotely sensed and site data." *International Journal of Remote Sensing* 23 (2002): 1125-1142.

Murthy, C. S., P. V. Paju, and K. V.S. Badrinath. "Classification of wheat crop with multi-temporal images: Performance of maximum likelihood and artificial neural networks." *International Journal of Remote Sensing* 24 (2003): 4871-4890.

Oh, D. *Radiometric Correction of Mobile Laser Scanning Intensity Data*. MSc thesis, International Institute for Geo-information Science and Earth Observation, The netherlands, 2010.

Pal, M. "Random forest classifier for remote sensing classification." *International Journal of Remote Sensing* 26, no. 1 (2005): 217-222.

Pal, M., and P. M. Mather. "An assessment of the effectiveness of decision tree methods for land cover classification." *Remote Sensing of Environment* 86 (2003): 554-565.

Pal, M., and P. M. Mather. "Some issues in the classification of DAIS hyperspectral data"." *International Journal of Remote Sensing* 27 (2006): 2895-2916.

Pal, M., and P. M. Mather. "Support vector machines for classification in remote sensing." *International Journal of Remote Sensing* 26, no. 5 (2005): 1007-1011.

Pebesma, E. J., and R. S. Bivand. "Classes and methods for spatial data in R." *R News* 5, no. 2 (2005): http://cran.r-project.org/doc/Rnews/.

Pellizzeri, T. M., P. Gamba, P. Lombardo, and F. Dell Acqua. "Multitemporal/multiband SAR classification of urban areas using spatial analysis: Statistical versus neural kernel-based approach." *IEEE Transactions on Geoscience and Remote Sensing* 41 (2003): 2338-2353.

Petropoulos, G. P., K. P. Vadrevu, G. Xanthopoulos, G. Karantounias, and M. Scholze. "A comparison of spectral angle mapper and artificial neural network classifiers combined with Landsat TM imagery analysis for obtaining burnt area mapping." *Sensors* 10 (2010): 1967-1985.

Platt, J. C. "Probabilistic outputs for Support Vector Machines and comparison to regularized likelihood methods." In *Advances in Kernel Methods Support Vector Learning*, edited by B. Schölkopf, C.J. C. Burges and A. J. Smola , 61-74. Cambridge, MA: MIT Press, 2000.

R Development Core Team. "R: A Language and Environment for Statistical Computing,Vienna, Austria." 2009. http://www.R-project.org.

Rabe, A., S. van der Linden, and P. Hostert. "Simplifying support vector machines for classification of hyperspectral imagery and selection of relevant features." *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*, 2010: 1-4.

Ran, Y. H., X. Li, L. Lu, and Z. Y. Li. "Large-scale land cover mapping with the integration of multi-source information based on the Dempster-Shafer theory." *International Journal of Geographical Information Science* 26, no. 1 (2012): 169-191.

Ratle, F., G. Camps-Valls, and J. Weston . "Semisupervised Neural Networks for efficient hyperspectral image classification." *IEEE Transactions on Geoscience and Remote Sensing* 48, no. 5 (2010): 2271-2282.

Rodr ́guez-Galiano, V. F., M. Chica-Olmo, F. Abarca-Hernandez, P.M. Atkinson, and C. Jeganathan. "Random forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture." *Remote Sensing of Environment* 121 (2012): 93-107.

Rottensteiner, F. "Automation of object extraction from LiDAR in urban areas." *2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2010.

Rottensteiner, F., and S. Clode. "Building and road extraction by Lidar and imagery." In *Topographic Laser Ranging and Scanning: Principles and Processing*, edited by J. Shan and C. K. Toth, 445-478. Boca Raton, FL: CRC Press, 2009.

Rottensteiner, F., J. Trinder, S. Clode, and K. Kubik. "Using the Dempster Shafer method for the fusion of LIDAR data and multispectral images for building detection." *Information Fusion* 6, no. 4 (2005): 283-300.

Salah, M., and J. Trinder. "Support vector machines based filtering of Lidar data: A grid cased method." *FIG Congress 2010*, 2010.

Samadzadegan, F., B. Bigdeli, and P. Ramzi . "Classification of LiDAR data based on multi-class SVM." *ISPRS Archives of The 2010 Canadian Geomatics Conference*

and Symposium of Commission I XXXVIII, no. 1 (2010a): 1-6 http://www.isprs.org/proceedings/XXXVIII/part1/03/03_01_Paper_185.pdf.

Samadzadegan, F., B. Bigdeli, and P. Ramzi. "A multiple classifier system for classification of LIDAR remote sensing data using multi-class SVM." In *Proceeding of Multiple Classifier Systems, 9th International Workshop*, edited by N. El Gayar, J. Kittler and F. Roli, 254-263. Verlag Berlin Heidelberg: Springer, 2010b.

Santana, L. M. "Landsat ETM+ image applications to extract information for environmental planning in a Colombian city." *International Journal of Remote Sensing* 28 (2007): 4225-4242.

Sarma, V., and X. Yuan. *Urban Surface Characterization Using LiDAR and Aerial Imagery.* Master thesis, University of North Texas, US, 2009.

Schmidt, K. S., A. K. Skidmore, E. H. Kloosterman, H. van Oosten, L. Kumar, and J. A. M. Janssen. "Mapping coastal vegetation using an expert system and hyperspectral imagery." *Photogrammetric Engineering and Remote Sensing* 70 (2004): 703-715.

Schneider, A., and C. E. WoodCock. "Compact, dispersed, fragmented, extensive? A comparison of urban growth in 25 global cities using remotely sensed data, pattern metrics and census information." *Urban Studies* 45 (2008): 659-692.

Schwalbe, E., H. G. Maas, and F. Seidel. "3D building model generation from airborne laser scanner data using 2D GIS data and orthogonal point cloud projections." *ISPRS WG III/3,III/4, V/Workshop "Laser Scanning 2005", Enschede, the Netherlands, September 12-14, 2005.*, 2005.

Secord, J., and A. Zakhor. "Tree detection in urban region using eerial Lidar and image data." *IEEE GeoScience and Remote Sensing Letters* 4, no. 2 (2007): 196-200.

Shah, C. A., M. K. Arora, and P. K. Varshney. "Unsupervised classification of hyperspectral data: An ICA mixture model based approach." *International Journal of Remote Sensing* 25 (2004): 481-487.

Shan, J., and A. Sampath. "Urban DEM generation from raw lidar data: A labelling algorithm and its performance." *Photogrammetric Engineering & Remote Sensing* 71, no. 2 (2005): 217-226.

Sithole, G., and G. Vosselman. "Automatic structure detection in a point-cloud of an urban landscape." *Remote Sensing and Data Fusion over Urban Areas, 2nd GRSS/ISPRS Joint Workshop*, 2003: 67-71.

Small, C. "Urban remote sensing: Global comparisons." *Architectural Design* 178 (2005): 18-23.

Smith, A. "Image segmentation scale parameter optimization and land cover classification using the Random Forest algorithm." *Journal of Spatial Science* 55, no. 1 (2010): 69-79.

Sohn, G., and I. Dowman. "Data fusion of high-resolution satellite imagery and Lidar data for automatic building extraction." *ISPRS Journal of Photogrammetry & Remote Sensing* 62 (2007): 43-63.

Starek, M. J., R. K. Vemula, K. C. Slatton, R. L. Shrestha, and W. E. Carter. "Shoreline based feature extraction and optimal feature selection for segmenting airborne LiDAR intensity images." *Proceeding of ICIP* 4 (2007): 369-372.

Stathakis, D. "How many hidden layers and nodes?" *International Journal of Remote Sensing* 30, no. 8 (2009): 233-2147.

Statistics Canada. "Niagara Falls, Ontario (Code3526043) (table). 2006 Community Profiles. 2006 Census. Statistics Canada Catalogue no. 92-591-XWE. Ottawa. Released March 13, 2007." *Niagara Falls, Ontario (Code3526043) (table), 2006 Community Profiles, 2006 Census, Statistics Canada Catalogue no. 92-591-XWE, Ottawa.* 2007. http://www12.statcan.ca/census-recensement/2006/dp-pd/prof/92-591/index.cfm?Lang=E (accessed October 9, 2011).

Stavrakoudis, D. G., G. N. Galidaki, I. Z. Gitas, and J. B. Theocharis. "A genetic fuzzy-rule-based classifier for land cover classification from hyperspectral imagery." *IEEE Transaction on GeoScience and Remote Sensing* 50, no. 148 (2012): 130-148.

Stefanov, W. L., and M. Netzband. "Assessment of ASTER land cover and MODIS NDVI data at multiple scales for ecological characterization of an urban center." *Remote Sensing of Environment* 99 (2005): 31-43.

Steinberg, D., and P. Colla. *CART -- Classification and Regression Trees.* San Diego, CA: Salford Systems, 1997.

Story, M., and R. Congalton. "Accuracy assessment: A user's perspective." *Photogrammetric Engineering and Remote Sensing* 52, no. 3 (1986): 397–399.

Stumpf, A., and N. Kerle. "Object-oriented mapping of landslides using random forests." *Remote Sensing of Environment* 115 (2011): 2564-2577.

Sveinsson, J. R., J. A. Benediktsson, and H. Aanaes. "Classification of pansharpened urban satellite images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5, no. 1 (2012): 281-297.

Swain, P. H., and S. M. Davis. *Remote Sensing: The Quantitative Approach.* New York, NY: McGraw Hill, 1978.

Thomas, N., C. Hendrix, and R. G. Congalton. "A comparison of urban mapping methods using high-resolution digital imagery." *Photogrammetric Engineering and Remote Sensing* 69 (2003): 963-972.

Tortora, R. "A note on sample size estimation for multinomial population." *The American Statistician* 32, no. 3 (1978): 100-102.

Trinder, J. C., and M. Salah. "Support vector machines: Optimization and validation for land cover mapping using aerial images and Lidar data." Sydney: 34th International Symposium on Remote Sensing of Environment, 2011.

Turner, B. L., E. F. Lambin, and A. Reenberg. "The emergence of land change science for global environment change and sustainability." *Processings of the National Academy of Sciences, USA* 104 (2007): 20666-20671.

Vapnik, V. *Estimation of Dependences Based on Empirical Data (in Russian).* Nauka, Moscow: pp.5165-5184 (English translation: Springer Verlag, New York, 1982), 1979.

Vatsavai, R. R., E. Bright, C. Varun, B. Budhendra, A. Cheriyadat, and J. Grasser. "Machine learning approaches for high-resolution urban land cover classification: A comparative study." *Proceedings of the 2nd International Conference on Computing for Geospatial Research & Applications*, 2011.

Verbeke, L. P. C., F. M. B. Vancoillie, and R. R. D. Wulf. "Reusing back-propagation artificial neural networks for land cover classification in tropical savannahs." *International Journal of Remote Sensing* 25 (2004): 2747-2771.

Vosselman, G., and S. Dijkman. "3D building model reconstruction from point clouds and ground plans." *International Archives of Photogrammetry and Remote Sensing* 34, no. 3-w4 (2001): 37-43.

Walter, V. "Object-based classification of remote sensing data for change detection." *ISPRS Journal of Photogrammetry & Remote Sensing* 58 (2004): 225-238.

Wang, L., G. S. Biging, and P. Gong. "Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery." *International Journal of Remote Sensing* 25 (2004): 5655-5668.

Wang, X., B. Waske, and J. A. Benediktsson. "Ensemble methods for spectral-spatial classification of urban hyperspectral data." *In Proceedings of IGARSS* 4 (2009): 944-947.

Wang, Y. Y., and J. Li. "Feature-selection ability of the decision-tree algorithm and the impact of feature-selection/extraction on decision-tree results based on hyperspectral data." *International Journal of Remote Sensing* 29, no. 10 (2008): 2993-3010.

Waske, B., and J. A. Benediktsson. "Fusion of support vector machines for classificationof multisensor data." *IEEE Transactions on Geoscience and Remote Sensing* 45, no. 12 (2007): 3858-3866.

Waske, B., and M. Braun. "Classifier ensembles for land cover mapping using multitemporal SAR imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (2009): 450-457.

Waske, B., and S. van der Linden. "Classifying multilevel imagery from SAR and optical sensors by decision fusion." *IEEE Transactions on Geoscience and Remote Sensing* 46, no. 5 (2008): 1457-1466.

Waske, B., J. A. Benediktsson, K. Arnason, and J. R. Sveinsson. "Mapping of hyperspectral AVIRIS data using machine-learning algorithms." *Canadian Journal of Remote Sensing* 35, no. S1 (2009a): S106-S116.

Waske, B., M. Chi, J. A. Benediktsson, S. van der Linden, and B. Koetz. "Algorithms and Applications for Land Cover Classification: A Review." In *Geospatial Technology for Earth Observation*, edited by D. Li, J. Shan and J. Gong, 203-233. Verlag Berlin Heidelberg: Springer, 2009b.

Waske, B., M. Fauvel, J. A. Benediktsson, and J. Chanussot. "Machine learning yechniques in remote sensing data analysis." In *Kernel Methods for Remote Sensing Data Analysis*, edited by Gustavo Camps-Valls and L. Bruzzone, 3-24. UK: Jonh Wiley&Sons Ltd, 2009c.

Watanachaturaporn, P., M. K. Arora, and P. K. Varshney. "Multisource classification using support vector machines: An empirical comparison with decision tree and neural network classifiers." *Photogrammetric Engineering & Remote Sensing* 74, no. 2 (2008): 239-246.

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

Weng, Q. "Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends." *Remote Sensing of Environment*, 2011.

Weng, Q., and D.A. Quattrochi . *Urban Remote Sensing.* Boca Raton, FL: Taylor&Francis Group, 2007.

Wentz, E. A., D. Nelson, A. Rahman, W. L. Stefanov, and S. S. Roy. "Expert system classification of urban land use/cover for Delhi, India." *International Journal of Remote Sensing* 29, no. 15 (2008): 4405-4427.

Weston, J., and C. Watkins. "Support vector machines for multi-class pattern recognition." *Proceedings of the Seventh European Symposium on Artificial Neural Networks*, 1999.

Wu, S. S., X. Qiu, L. Usery, and L. Wang. "Using geometrical, textural, and contextual information of land parcels for classification of detailed urban land use." *Annals of the Association of American Geographers* 99, no. 1 (2009): 76-98.

Wu, T.-F., C.-J. Lin, and R.-C. Weng. "Probability estimates for multi-class classification by pairwise coupling." In *Advances in Neural Information Processing Systems 16*, edited by S. Thrun, L. K. Saul and B. Schölkopf. Cambridge, MA: MIT Press, 2004.

Yang, H., Q. Du, and G. Chen. "Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* Early Access Articles (2012): 1-11.

Yang, X.. "Parameterizing support vector machines for land cover classification." *Photogrammetric Engineering & Remote Sensing* 77, no. 1 (2011a): 27-37.

Yang, X.. *Urban Remote Sensing.* John Wiley& Sons Ltd., 2011b.

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

Zhan, Q., and L. Yu. "Objects classification from laser scanning data based on multi-class support vector machine." *2011 International Conference on Remote Sensing, Environment and Transportation Engineering (RSETE)*, 2011: 520-523.

Zhang, J. "Multi-source remote sensing data fusion: status and trends." *International Journal of Image and Data Fusion* 1, no. 1 (2010): 5-24.

Zhang, L., L. Zhang, and D. Tao. "On combining multiple features for hyperspectral remote sensing image classification." *IEEE Transaction on Geoscience and Remote Sensing* 50, no. 3 (2012): 879-893.

Zhong, Y., and L. Zhang. "An adaptive artificial immune network for supervised classification of multi-/hyperspectral remote sensing imagery." *IEEE Transaction on Geoscience and Remote Sensing* 50, no. 3 (2012): 894-909.

Zhou, W., and A. Troy. "An object-oriented approach for analysing and characterizing urban landscape at the parcel level." *International Journal of Remote Sensing* 29, no. 11 (2008): 3119-3135.

Zhu, Z., C. E. Woodcock, J. Rogan, and J. Kellndorfer. "Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data." *Remote Sensing of Environment, doi:10.1016/j.rse.2011.07.020*, 2011.

Appendix A

Feature Images

This appendix provides illustrations for all features used in this study (see Table 0-1).

Figure 1 “Red”

Figure 2 “Green”

Figure 3 “Blue”

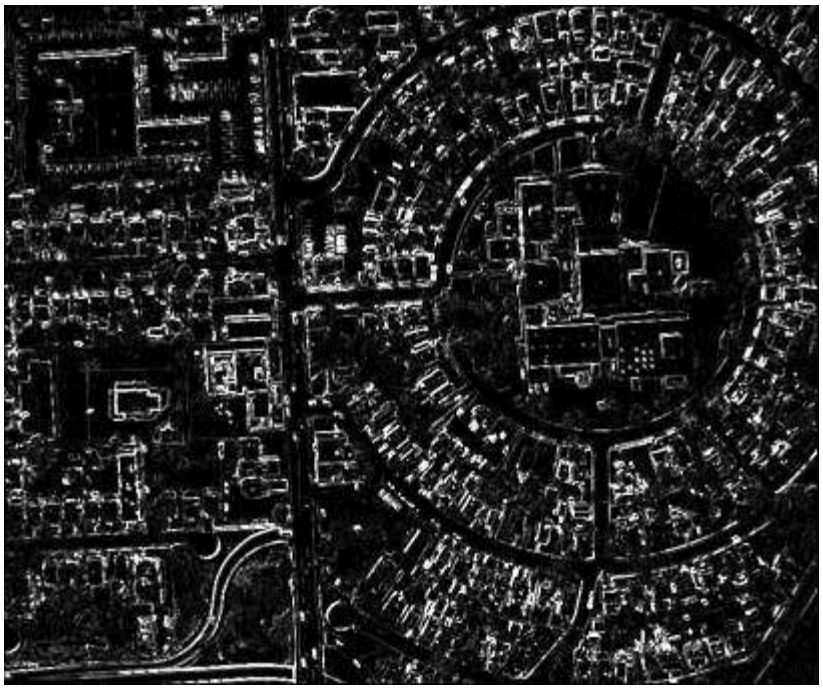


Figure 4 “Red_Contrast”

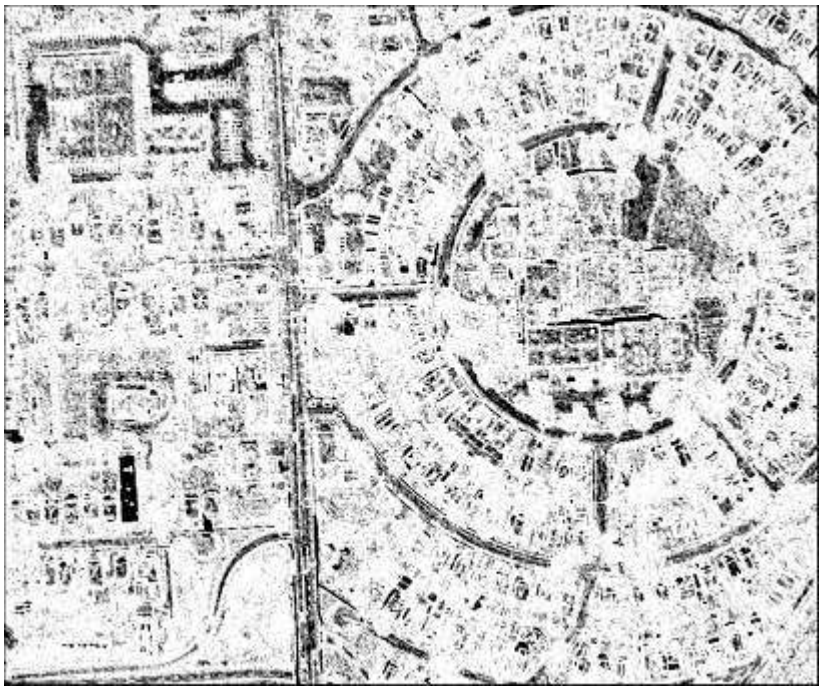


Figure 5 “Red_Entropy”

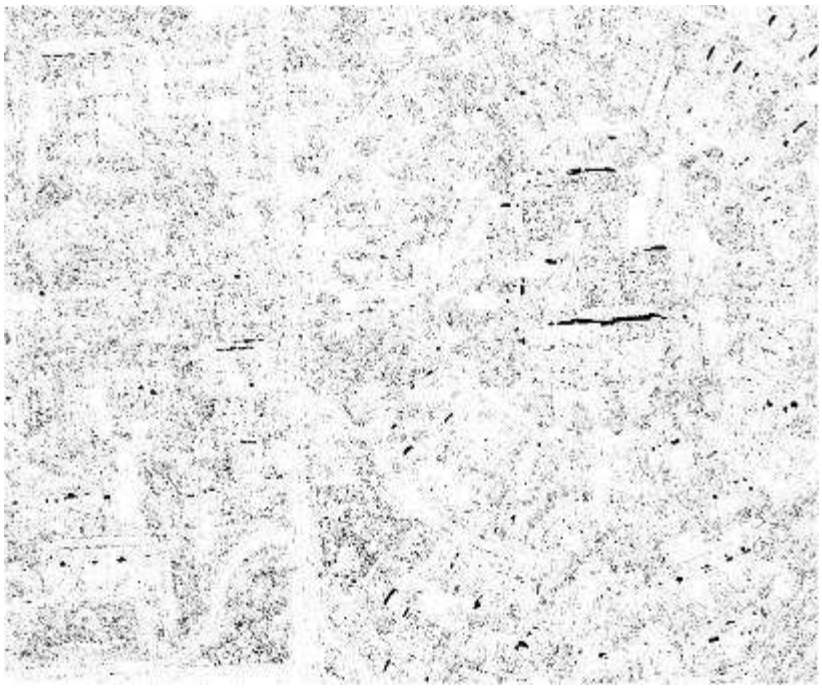


Figure 6 “Red_Correlation”

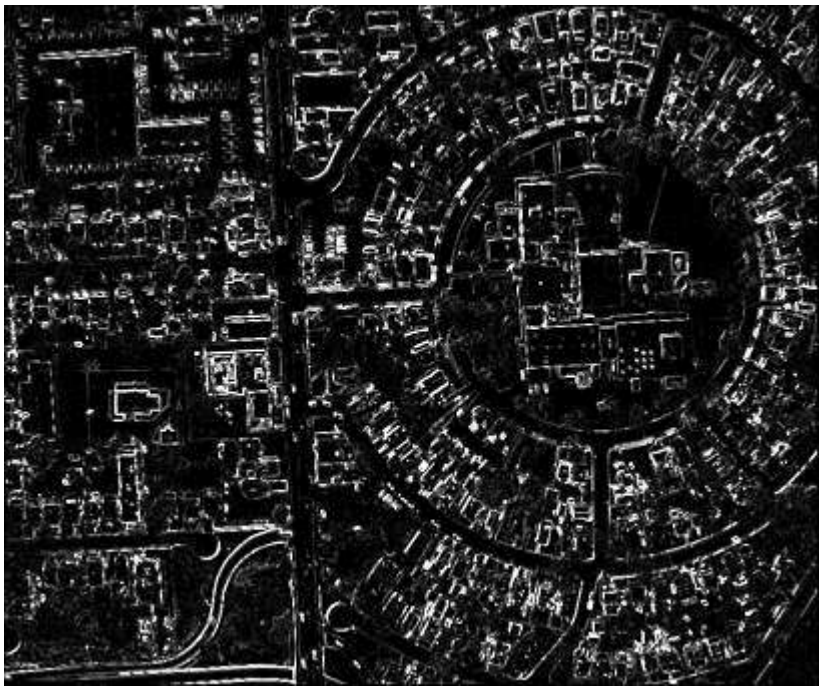


Figure 7 “Green_Contrast”

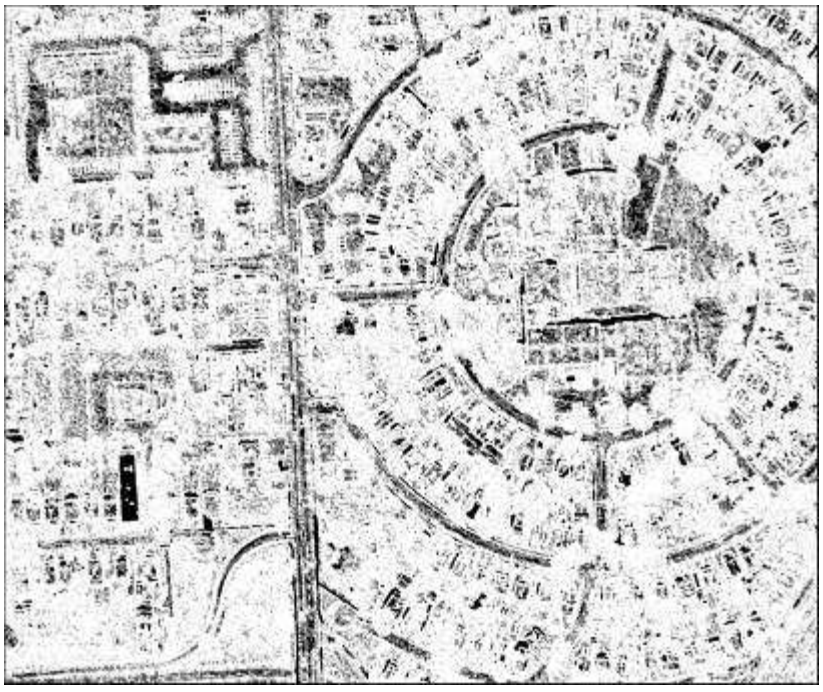


Figure 8 “Green_Entropy”

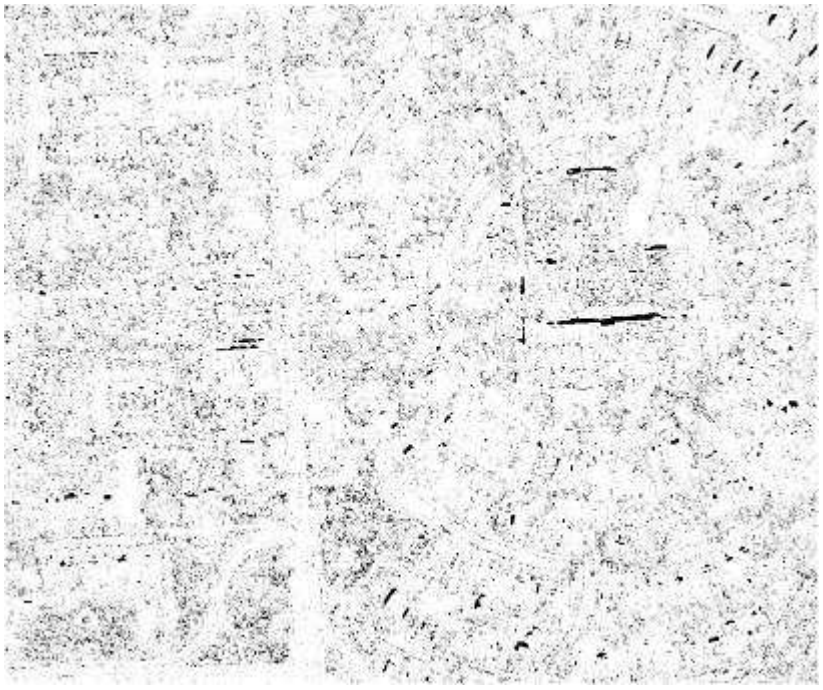


Figure 9 “Green_Correlation”

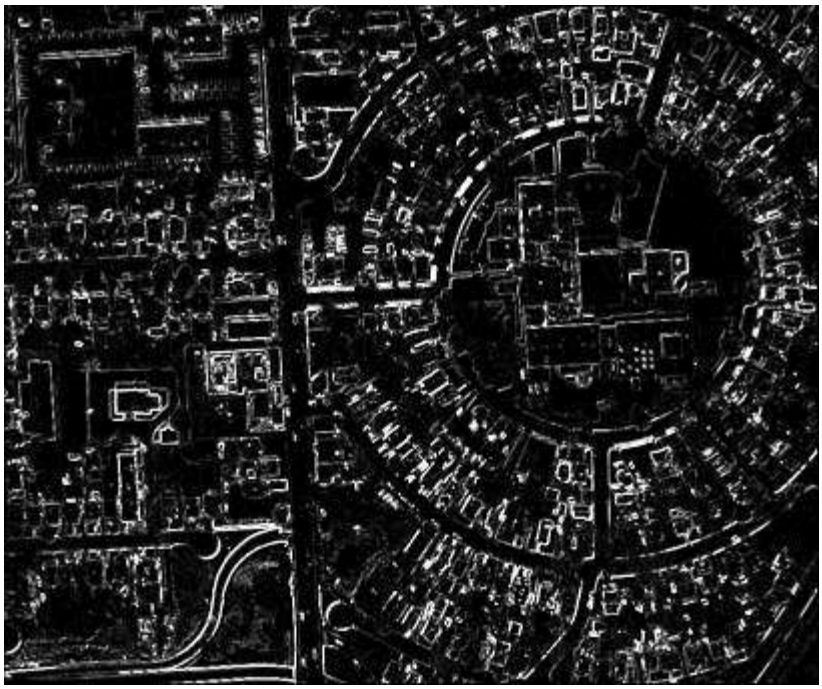


Figure 10 “Blue_Contrast”

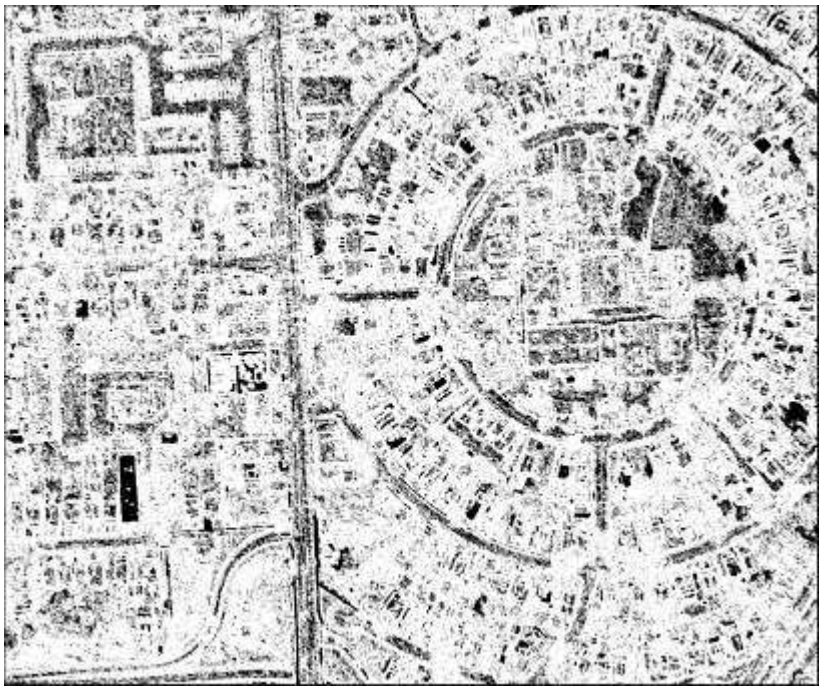


Figure 11 “Blue_Entropy”

Figure 12 “Blue_Correlation”

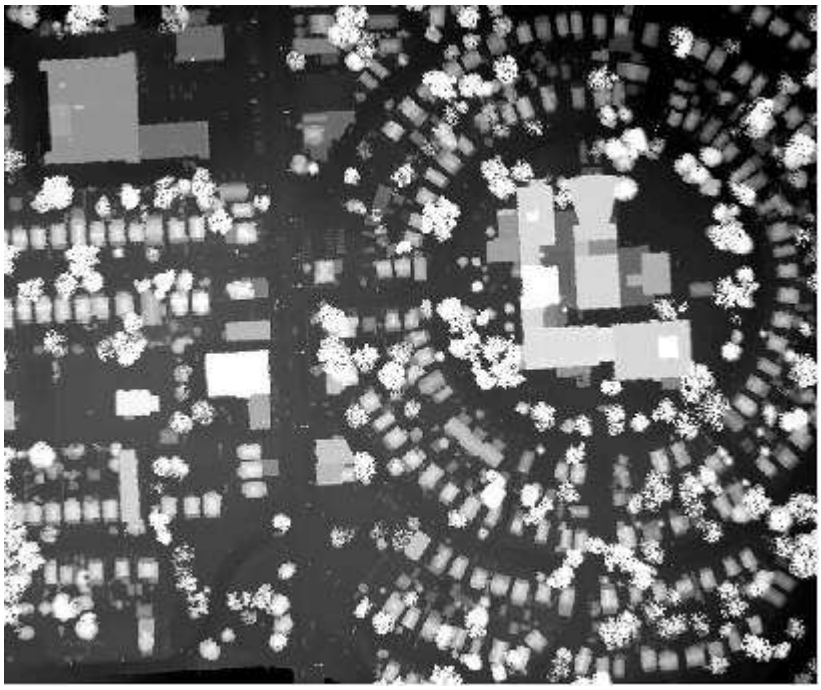


Figure 13 “DSM”

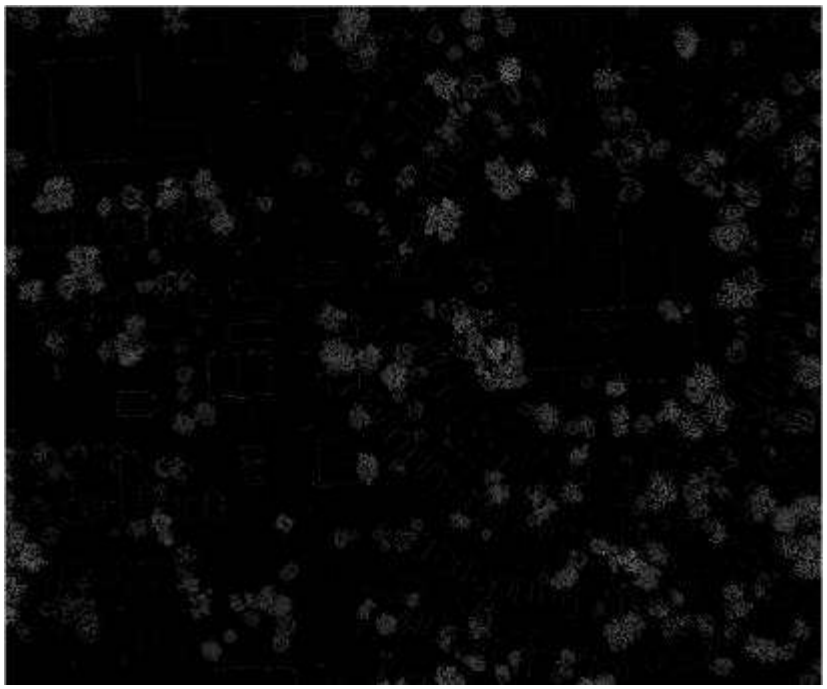


Figure 14 “fr-lr”

Figure 15 “HDiff”

Figure 16 “HVar”

Figure 17 “LiDAR Intensity”

Figure 18 “DSM_Contrast”

Figure 19 “DSM_Entropy”

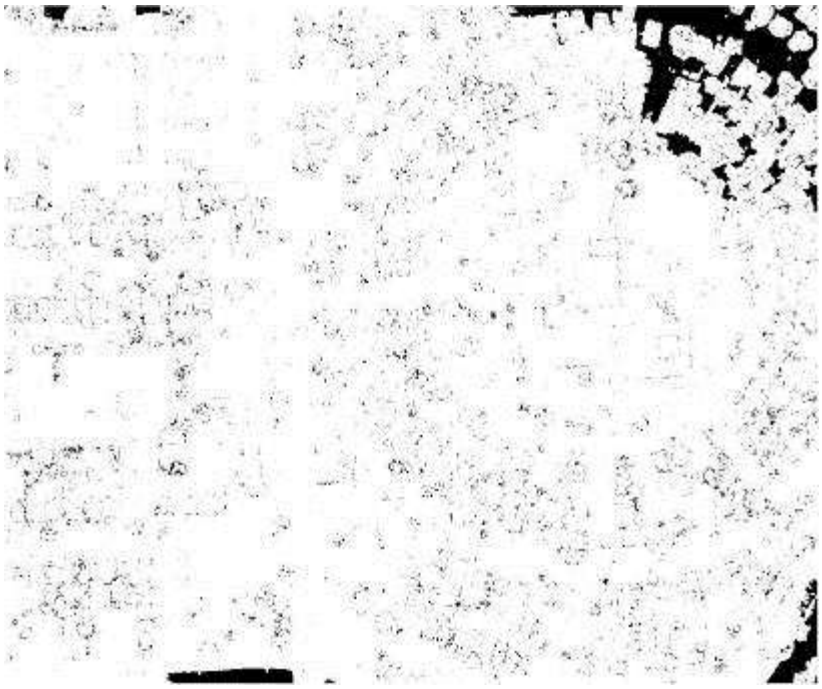


Figure 20 “DSM_Correlation”

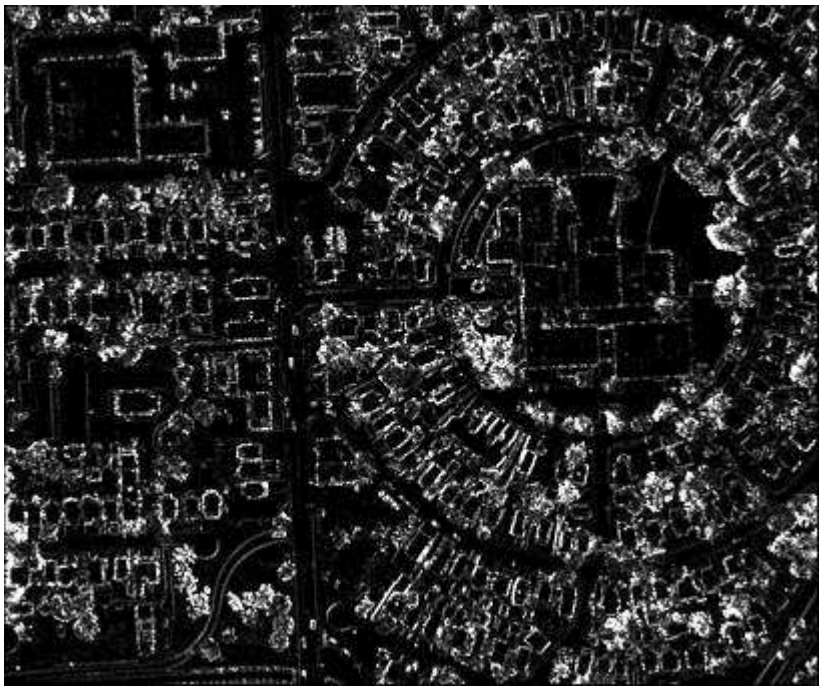


Figure 21 “Intensity_Contrast”

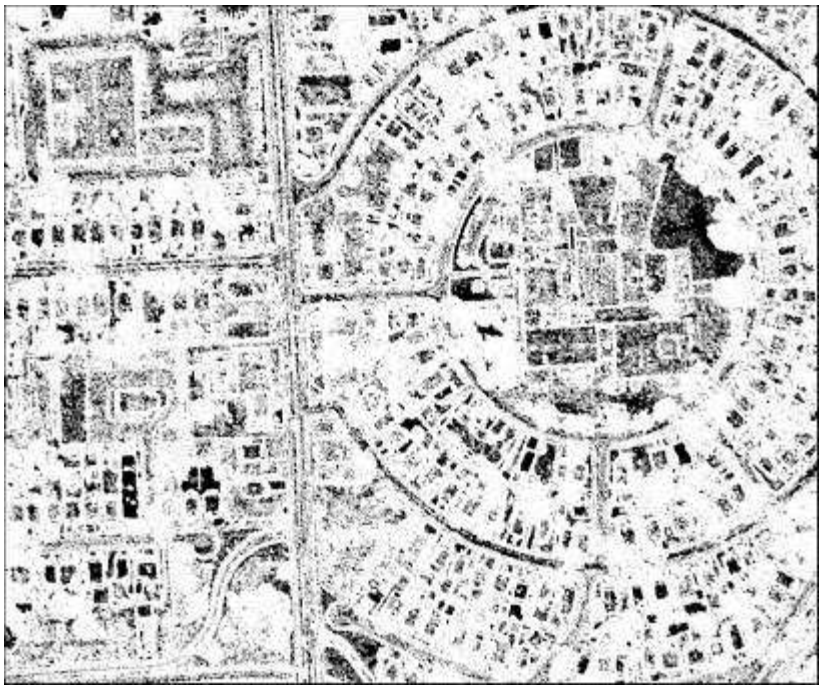


Figure 22 “Intensity_Entropy”

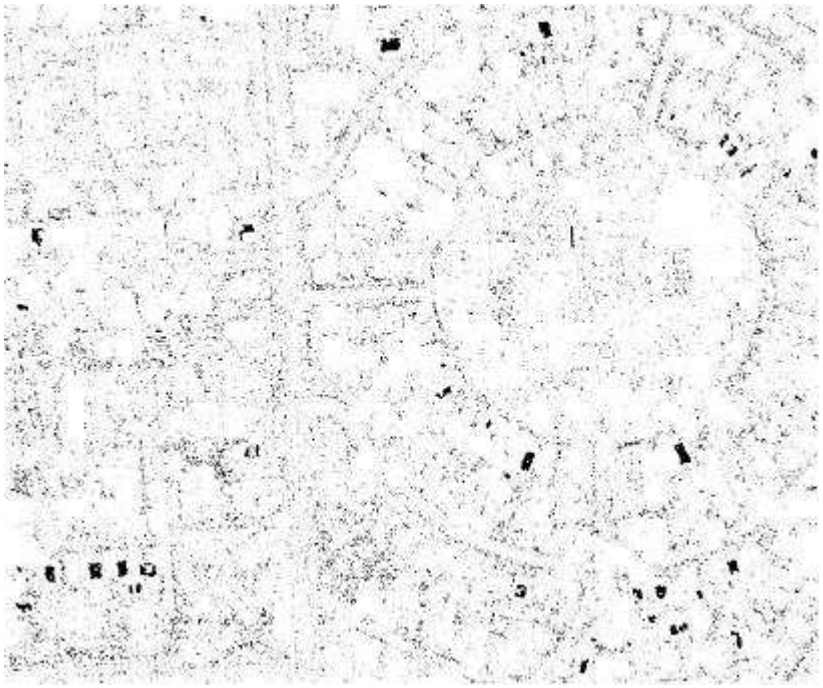


Figure 23 “Intensity_Correlation”

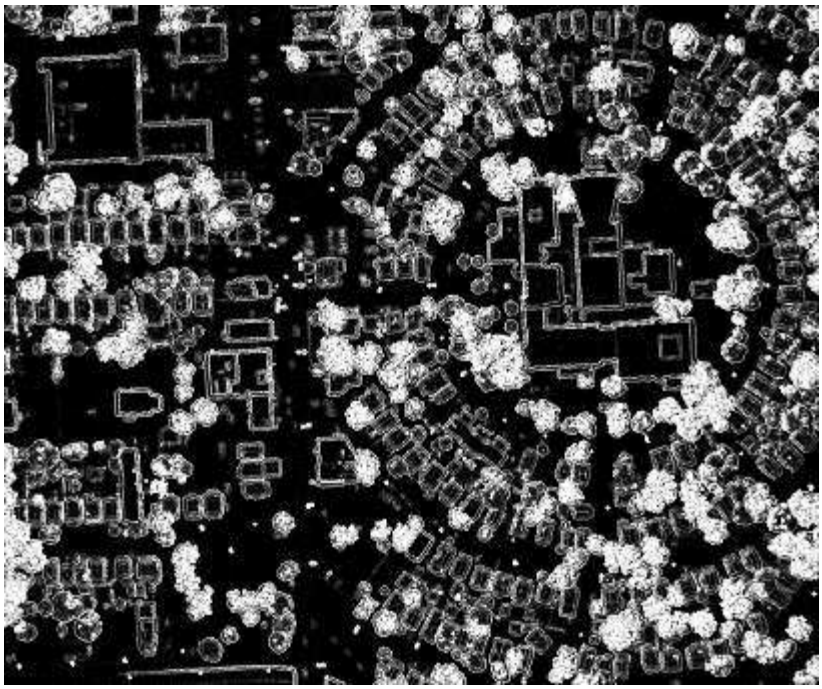


Figure 24 “Eigen1”

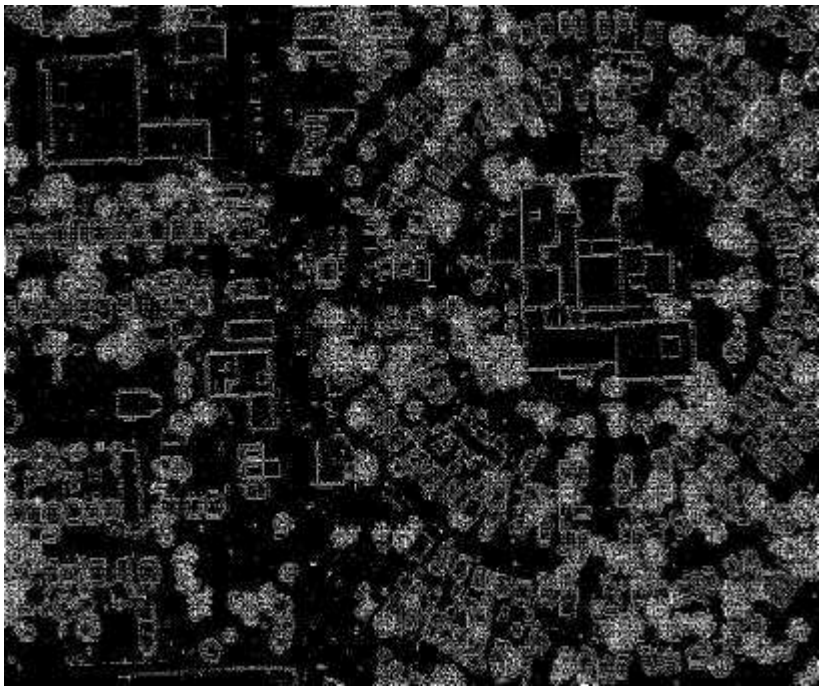


Figure 25 “Eigen 2”

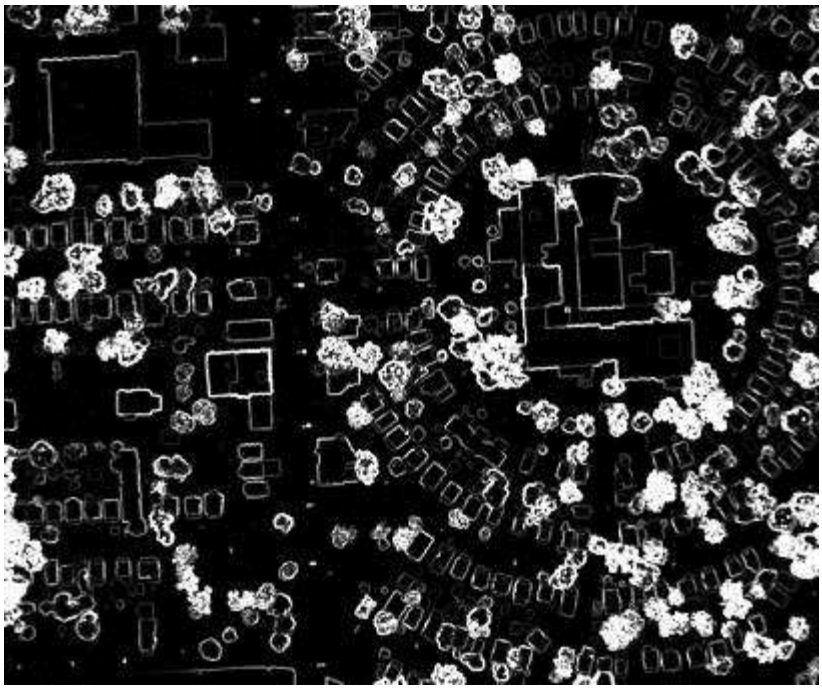


Figure 26 “Eigen3”

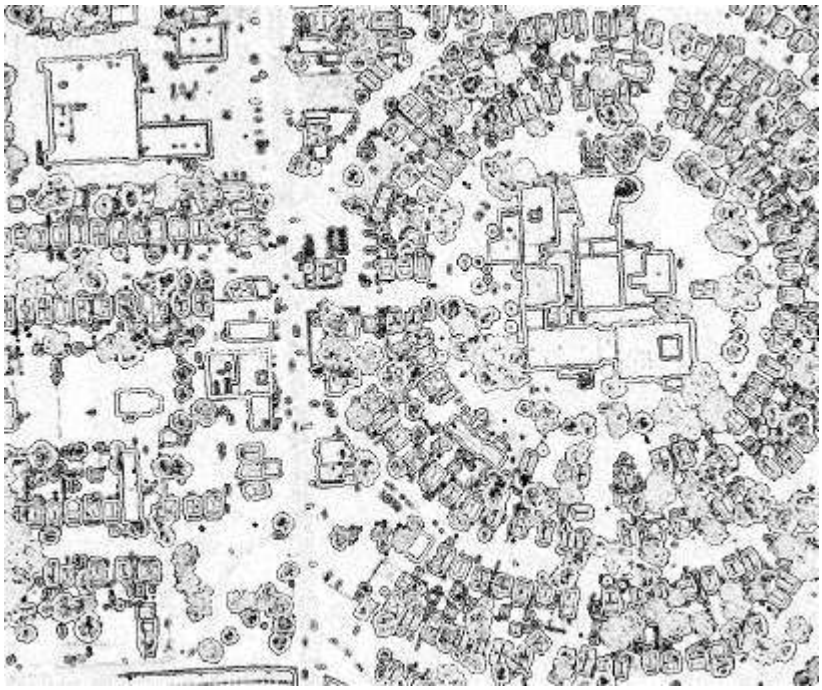


Figure 27 “Anisotropy”

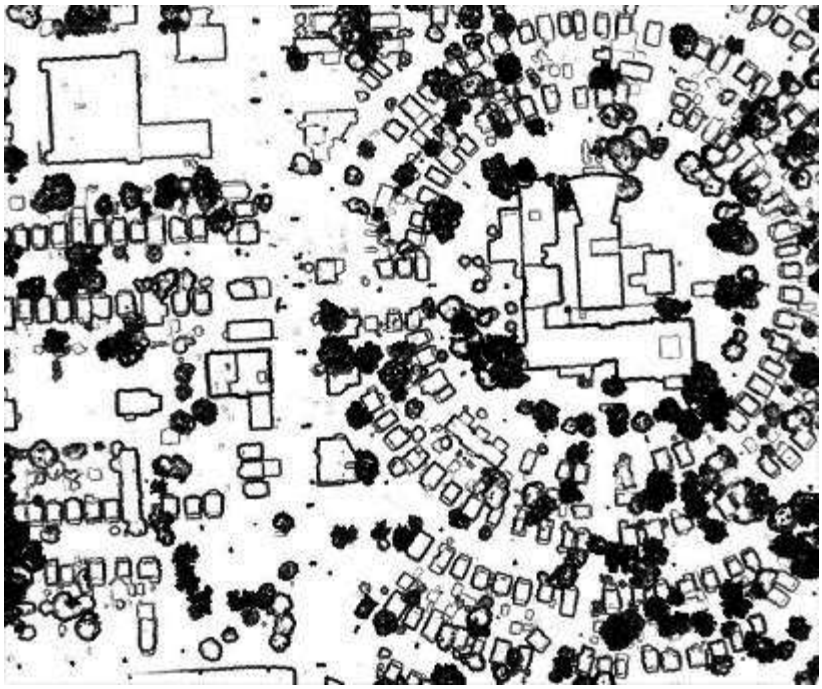


Figure 28 “Planarity”

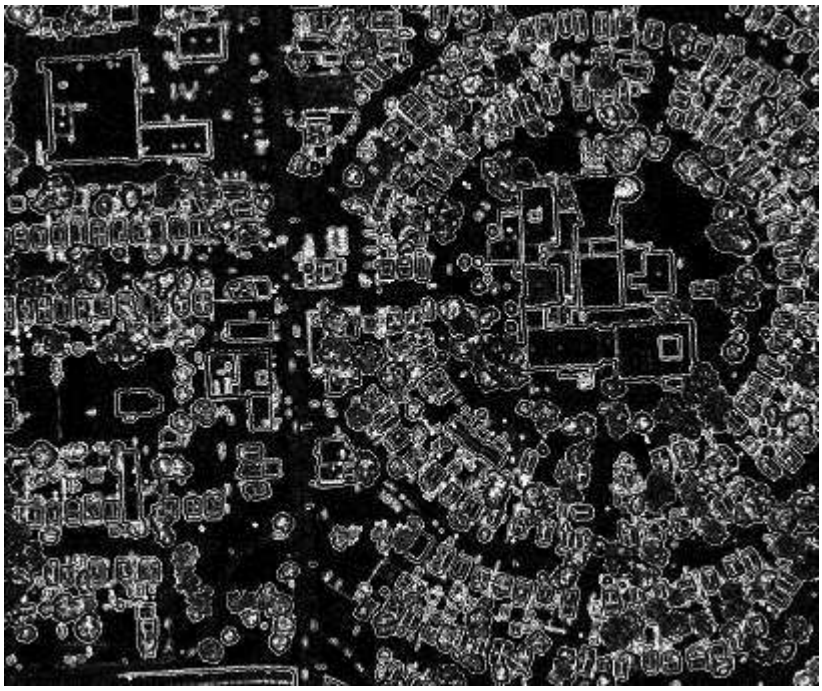


Figure 29 “Sphericity”

Figure 30 “Linearity”

Figure 31 “LiDAR_TVI”