International Journal of Remote Sensing, 2013
Vol. 34, No. 14, 5166–5186, http://dx.doi.org/10.1080/01431161.2013.788261

Taylor & Francis
Taylor & Francis Group

Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests

Haiyan Guan[a], Jonathan Li[a,b]*, Michael Chapman[c], Fei Deng[d], Zheng Ji[e], and Xu Yang[f]

[a]*Department of Geography & Environmental Management, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1;* [b]*School of Information Science and Engineering, Xiamen University, Xiamen, FJ 3610005, China;* [c]*Department of Civil Engineering, Ryerson University, Toronto, Ontario, Canada M5B 2K3;* [d]*School of Geodesy & Geomatics, Wuhan University, Wuhan 430079, Hubei, China;* [e]*School of Remote Sensing & Information Engineering, Wuhan University, Wuhan 430079, Hubei, China;* [f]*Unit 63961, Beijing 100012, China*

(*Received 4 July 2012; accepted 14 March 2013*)

Using high-spatial-resolution multispectral imagery alone is insufficient for achieving highly accurate and reliable thematic mapping of urban areas. Integration of lidar-derived elevation information into image classification can considerably improve classification results. Additionally, traditional pixel-based classifiers have some limitations in regard to certain landscape and data types. In this study, we take advantage of current advances in object-based image analysis and machine learning algorithms to reduce manual image interpretation and automate feature selection in a classification process. A sequence of image segmentation, feature selection, and object classification is developed and tested by the data sets in two study areas (Mannheim, Germany and Niagara Falls, Canada). First, to improve the quality of segmentation, a range image of lidar data is incorporated in an image segmentation process. Among features derived from lidar data and aerial imagery, the random forest, a robust ensemble classifier, is then used to identify the best features using iterative feature elimination. On the condition that the number of samples is at least two or three times the number of features, a segmentation scale factor has no particular effect on the selected features or classification accuracies. The results of the two study areas demonstrate that the presented object-based classification method, compared with the pixel-based classification, improves by 0.02 and 0.05 in kappa statistics, and by 3.9% and 4.5% in overall accuracy, respectively.

1. Introduction

Since airborne laser scanning or light detection and ranging (lidar) systems have become widely used for the acquisition of three-dimensional (3D) data, a variety of lidar-based approaches to urban thematic mapping have been developed over the last two decades (Brenner 2010). In the early period of lidar techniques, lidar data, as a single data source, were researched in a number of applications, including building detection (Maas and Vosselman 1999; Vögtle and Steinle 2003; Li and Guan 2011) and tree inventory (Zimble et al. 2003; Brandtberg 2007). Although use of lidar data alone has advantages for urban object detection, some limitations exist, such as data gaps caused by wet buildings and roads, and recognition problems regarding buildings and trees. Therefore, in recent years,

*Corresponding author. Email: junli@xmu.edu.cn; junli@uwaterloo.ca

research on urban mapping has involved the integration of lidar data with data sources including IKONOS (Shan and Lee 2003), QuickBird (Chen et al. 2009), SPOT-5 (Alonso and Malpica 2010), GeoEye (Yu et al. 2011), and aerial images (Huang et al. 2008; Khoshelham et al. 2010; Guan et al. 2013). The philosophy behind the integration of lidar and optical imagery is that the strengths of one data type can compensate for the weaknesses of others. For example, being short of spectral information, lidar data have high classification confusion between man-made and natural objects, whereas multispectral image data have increasing classification confusion between spectrally identical objects in intricate urban landscapes.

In the field of remote sensing, studies on urban thematic mapping (or urban land-use/land-cover classification) have become an essential component in analysing the interactions between human activities and physically environmental changes. In most urban areas, the focus is on the following four main objects: buildings, trees, grass, and bare ground (Rottensteiner et al. 2005; Huang et al. 2008; Guo et al. 2011; Guan, Li, and Chapman 2011). Buildings and bare ground, as dominant objects in urban areas, characterize important information relating to urban landscape and human activities. Undoubtedly, they are widely used in many applications, including (1) urban solar energy potential mapping and collection, (2) 3D city modelling, (3) geographical information system (GIS) database updating, (4) environmental planning, (5) engineering surveying, and (6) topographical mapping. Moreover green infrastructure, an interconnected network of green space, conserves natural ecosystem values and functions and provides associated benefits to human populations (Benedict and McMahon 2002). As two important elements of green infrastructure, Tree and Grass are primarily chosen for analysis of the quantity and quality of green space, which includes energy exchange and hydrological modelling, heat island effect, 3D modelling, and climate change. Therefore, understanding the spatial distribution of human activities and physical environment at various levels plays a critical role in sustainable development.

A number of classification methods for urban thematic mapping have been well developed, ranging from (1) unsupervised mean-shift (Melzer 2007) and ISODATA (Germaine and Huang 2011) algorithms, through (2) supervised classification methods that include traditional maximum likelihood (Haala and Walter 1999), a recently-vigorous machine-learning-encompassing a support vector machine (Secord and Zakhor 2007; Mallet, Bretar, and Soergel 2008; Guan et al. 2011) and random forests (Chehata, Guo, and Mallet 2009; Guo et al. 2011), to (3) object-based classification or object-oriented classification, or object-based image analysis (Guan et al. 2011; Yu et al. 2011). A traditional pixel-based classification is often inadequate for high-resolution multispectral images in the complex urban environment (Zhou and Troy 2008), as it suffers from the well-known salt-and-pepper effect. To overcome this situation, several object-based classification methods have been proposed, in which neighbouring pixels are grouped together into image objects in a process called segmentation. These image objects are then classified according to their attributes, such as colour/value, shape, texture, or context (Lehrbass and Wang 2012). Owing to improvements in image segmentation, it has been claimed that object-based classification methods can classify land-cover/land-use efficiently (Duveiller et al. 2008). Since 2000, there has been a sharp increase in the popularity of object-based classification methods as opposed to pixel-based classification methods (Gamanya, De Maeyer, and De Dapper 2009; Blaschke 2010).

Feature selection is a commonly used process in land-use/land-cover classification, wherein a subset of features available from the data is selected for applying a classifier. In this way, the selected subset contains features that most contribute to classification.

In the context of remote sensing, feature selection is generally defined as a task to remove irrelevant and/or redundant features. Although a plethora of features can be extracted from both lidar point clouds and optical imagery, how to choose features for a given set of classes is an open problem for effective object recognition. It is well known that the subjective selection of features affects the quality of classification accuracy. Even for widely acclaimed object- and knowledge-oriented classifiers, it is difficult to decide which descriptive features are truly significant. Furthermore, most classifiers are limited to different types of input data and diverse environmental conditions. Since there is no easy way to decide the optimal number of features in advance, one can try a greedy feature selection until an acceptable level of accuracy is reached. Pineda-Bautista, Carrasco-Ochoa, and Martínez-Trinidad (2011) discussed two ways to eliminate redundant and irrelevant features: traditional feature selection for all classes and class-specific feature selection. In contrast with traditional feature selection, which selects a single feature for discriminating all classes, the class-specific feature selection algorithm selects a subset of features for each class. In most cases, features vary with landscape type for each expected class, but current feature selection algorithms have also been proposed for individual classifiers. Thus, a general framework for selecting optimal features is required.

Ensemble learning algorithms (e.g. bagging and boosting) have received increasing interest because a set of classifiers has a better classification performance than a single classifier (Breiman 1996). A random forest can be considered as an improved version of bagging. Compared with bagging and boosting, a random forest is characterized by computational efficiency, robustness to outliers and noises, and useful internal estimates of error, strength, correlation, and variable importance. In the remote-sensing domain, random forests have achieved promising classification accuracy for hyperspectral (Wang, Waske, and Benediktsson 2009), multispectral (Stumpf and Kerle 2011), and multisource data (Gislason, Benediktsson, and Sveinsson 2006). Due to the classification complexity of multisource data, commonly used parametric classification methods are inappropriate. The random forest, as a nonparametric classification algorithm, should be of great interest for multisource data by providing an estimate of an individual variable importance index. Therefore, it is of importance to investigate the influence of each predictive feature for selecting the best features when lidar data and aerial imagery are used in a classification model.

This article is organized as follows. Section 2 describes two study sites and data; in Section 3, the proposed object-based classification method for urban mapping is addressed by lidar and image feature extraction, image segmentation, random forest-based class-specific feature selection, and accuracy assessment. The experimental results are presented and discussed in Section 4. Finally, Section 5 concludes the article and suggests future research directions.

2. Study areas and data

2.1. Test data set 1: Mannheim, Germany

Mannheim, a city in southwestern Germany, is located at the confluence of the River Rhine and the River Neckar in the northwest corner of the state of Baden-Württemberg, and is unusual among German cities in that the streets and avenues in the central area are laid out in a grid pattern, just like most North American cities and towns. The data set represents a dense urban area with relatively flat elevation, ranging from approximately 89.83 to 159.71 m. It contains variously sized buildings of different orientations, as well as trees and grass interspersed among buildings. Laser scanning data covering the central area were

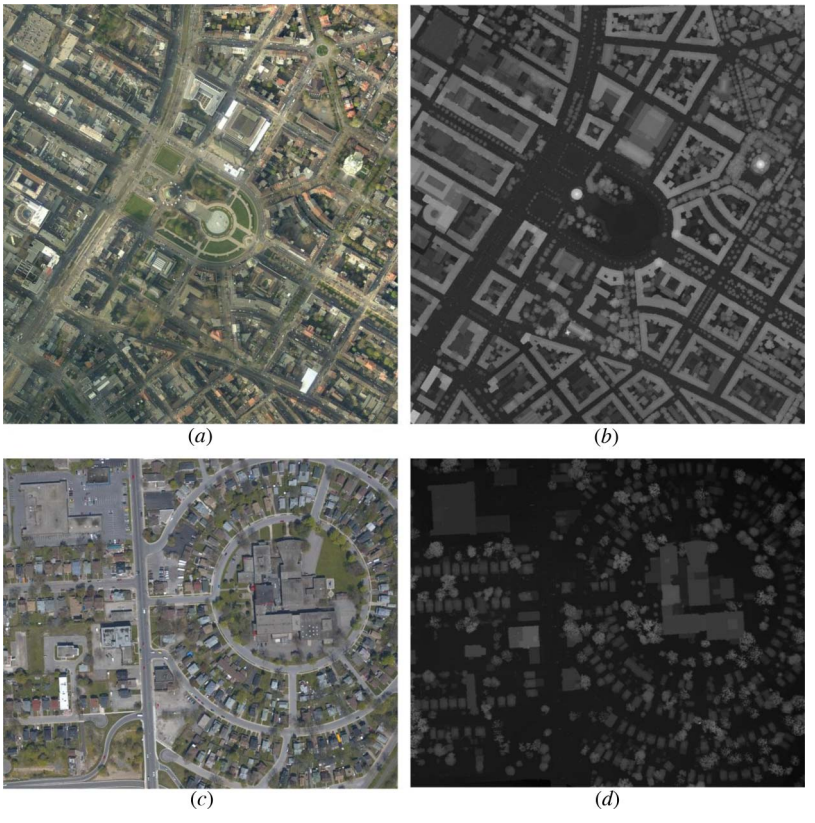


Figure 1. Test data sets: (*a*) colour orthoimage and (*b*) lidar range image of area in Mannheim (Germany); (*c*) colour orthoimage, and (*d*) lidar range image of area in Niagara Falls (Canada).

acquired in 2004 by a Falcon II sensor – the fibre-based system from TopoSys® GmbH (Biberach, Germany). The airplane flew at an average height of 1200 m above mean sea level, with a camera on board for 0.5 m resolution colour aerial photographs. The average point density and point spacing within the test site is about 4 points/m^2 and 0.5 m, respectively. The lidar data set records both range (first and last returns) and intensity information of the laser pulse, as shown in Figures 1(*a*) and (*b*).

2.2. Test data set 2: Niagara Falls, Canada

Niagara Falls is a city located on the Niagara River in the Golden Horseshoe region of Southern Ontario, Canada. There are a school, a shopping plaza, and more than 300 residential and commercial buildings in this study area. The study area and its vicinity are relative flat, with elevations ranging from approximately 148.71 to 178.11 m. 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 lidar data, containing first and last returns and intensity information, were acquired in 2004 by an Optech ALTM 3100 system at an average height of 1190 m above mean sea level, with a DSS 301 SN0039 camera on board for 0.5 m resolution colour aerial photographs, as shown in Figures 1(*c*) and (*d*). The horizontal and vertical accuracies of lidar data are

0.6 and 0.15 m (1 standard deviation), respectively. Similar to the Mannheim data set, the average point density and point spacing within this study site is about 4 points/m^2 and 0.5 m, respectively.

3. Method

Similar to a pixel-based classification method, the proposed object-based classification method uses all possible spectral and non-spectral bands as inputs. However, the difference between them is that each pixel in the former is identified separately, whereas in the latter, all pixels belonging to an object are grouped or clustered together for object identification (Walter 2004).

3.1. Lidar and image features

The classification scheme described is based on the integration of lidar data and multispectral aerial imagery, because the two types of data are complementary. Aerial imagery provides high-resolution and multispectral information in the visible range of the spectrum, while lidar data provide accurate geometric information and intensity in the near-infrared range of the spectrum. Multi-return characteristics of lidar systems also offer unique penetration information about vegetation. To combine lidar and multispectral data, we transformed lidar point clouds into a 2D range image. An overview of features used in this study is listed in Table 1, including spectral-based and lidar height/intensity-based features.

Spectral-based features. The RGB bands were used as three individual spectral features after a process of low-pass filtering or smoothing. Besides spectral information of an image region, the spatial properties, relationships between grey levels in neighbouring pixels which contribute to the overall appearance of the image, should be taken into account. The grey-level co-occurrence matrix (GLCM) proposed by Julesz (1962) and later by Haralick, Shanmugam, and Dinstein (1973) is considered to be one of the most popular methods to measure texture.

Lidar-based geometric features. Although the 2D lidar range image was used in the land-cover classification scheme described, lidar height-based features were directly calculated from original 3D point clouds in a given spherical neighbourhood. Mainly determined by point density, the given sphere is required to contain at least six points.

Lidar-based intensity features. In Jutzi and Gross (2009), intensity information is the physical power of incoming echoes, considered as the synonym for amplitude, reflectance, or energy in the terminology of laser scanning. As a consequence, leveraging intensity shows promise for recognizing objects in urban scenes. Similarly, GLCM was used for the intensity image to obtain eight textural measures in this study.

3.2. Segmentation

In this study, multiresolution image segmentation (MRIS) embedded in eCognition Developer software (Trimble®, Munich, Germany) was used to obtain a series of non-overlapping segments. MRIS, a commonly used algorithm in Earth Science, is a region-growing segmentation algorithm that starts from seed points and groups their adjacent

Table 1. Overview of features used for this study.

		Test features	Description	Numbers of metrics
Image	Spectral information	Spectral bands (R, G, and B)	A Perona-Malik diffusion method is used to remove random noises caused by external factors such as the atmospheric effect and instrumentation	3
	Texture	GLCM measures: contrast (*Cont.*), dissimilarity (*Diss.*), homogeneity (*Homo.*), entropy (*Ent.*), mean (*Mean*), variance (*Var.*), second-moment (*S-M*), correlation (*Corr.*)	Grey level co-occurrence matrix (GLCM) is calculated to measure texture	8*3 = 24
Lidar	Height-based information	H_{FL}: height difference between first and last echoes, $H_{FL} = H_{First\text{-}echo} - H_{Last\text{-}echo}$	It represents penetrability that describes how well the laser beam from the lidar system can penetrate through objects to the ground. This feature will help distinguish high-rise penetrable vegetation (Huang et al. 2008)	1
		NHFL: normalized height difference, $NH_{FL} = (H_{First\text{-}echo} - H_{Last\text{-}echo})/(H_{First\text{-}echo} + H_{Last\text{-}echo})$	The lidar-based vegetation index with respect to the first and last echoes will highlight the vegetation (Arefi, Hahn, and Lindenberger 2003)	1
		nDSM: normalized digital model derived by subtracting a digital terrain model (DTM) from DSM	This feature will help distinguish elevated objects from the ground or near-ground objects (Haala and Walter 1999)	1
		ΔH: local height variation representing the absolute distance between the maximum and minimum height values in a given sphere	This feature will assist in discriminating ground and non-ground objects (Ke, Quackenbush, and Im 2010)	1
		N_H: Normal variation	The average dot product of each normal with other normals within a given size of searching window. It has been experimentally set to 5 m by 5 m	1

(*Continued*)

Table 1. (*Continued*).

		Test features		Description	Numbers of metrics
		L_C: Local curvature		This feature is generated by Laplacian operator in a window of 5 m by 5 m	1
		Planar features: in a local neighbourhood, a plane is used to derive vector-related features	N_P: deviation angle of the plane normal vector from the vertical direction	This feature will assist in discriminating the ground with small values of deviation angles	1
			ΔZ: distance from the current point to the local estimated plane	This feature reflects the local height variation used for discriminating the ground from non-ground objects	1
		Eigenvalue-related features	E_A: anisotropy	Eigenvalue-related features are defined as a group of spatial features of each point by calculating a variance-covariance matrix of its neighbours. It is another auxiliary indicator for distinguishing planes, edges, corners, and volumes (Chehata, Guo, and Mallet 2009)	1
			E_P: planarity		1
			E_S: shpericity		1
			E_L: linearity		1
	Intensity texture	GLCM (*Cont., Diss., Homo., Ent., Mean, Var., S-M, Corr.*)		Similarly, GLCM is used for the intensity image to obtain eight textural measures	8
Combined features		Lidar-NDVI: Lidar NDVI $= \sqrt{\frac{\mathrm{NIR}-R}{\mathrm{NIR}+R} + 0.5}$		This feature will separate natural objects from man-made objects (Huang et al. 2008)	1

pixels according to a criterion of homogeneity, based on a user-defined threshold of segmentation scale (Benz et al. 2004). The segmentation scale is specified to directly control the sizes of segments. The greater the user-defined scale parameter, the larger the average size of segments. Additionally, the defining segments can be adjusted by relative weighting parameters of colour and shape. However, it is somewhat difficult to segment and classify high-spatial-resolution aerial imagery due to high spectral heterogeneity within classes. To overcome this issue, a lidar range image used as an additional band was integrated into the segmentation process. All mixed bands (blue, green, red, and grey values of the lidar range image) were equally weighted. The fused image was then partitioned into non-overlapping segments consisting of groups of relatively homogeneous pixels. To evaluate the impact of segmentation scales on the feature selection and class separability, image segmentation was performed at 11 different scale parameter settings (10, 20, 30, 40, 50, 55, 60, 70, 80, 90, and 100). Segmentation results generally vary with the data used and the scenes processed. It is reasonable to assess and compare segmentation results obtained at different scales for the same scene rather than at the same scale for different scenes. The colour criterion was given a default weight of 0.9, while the shape was assigned with the remaining weight of 0.1. These parameters were determined by visual interpolation of the image segmentation results, where segments were considered to be internally homogenous. After image segmentation, geometric and spectral features from lidar data and aerial imagery were calculated for each segment.

3.3. Random forests

The random forest classifier is an ensemble learning technique developed by Breiman (2001) based on a combination of a large set of decision trees, classification and regression trees (CART). Each tree is trained by selecting a random set of variables and a random sample drawn from the training data set. The training data are sampled with replacements to create a data set. This technique is often referred to as 'bootstrapping'. Based on the 'bootstrapping' technique, two-thirds of the training data, termed *inbag* data, are used to construct the tree, and the remaining one-third, known as the *out-of-bag* (OOB) data, are used to test the constructed tree for internal evaluation of its performance. The average misclassification over all trees is known as the OOB error estimate. The OOB error estimate is unbiased in predicting the performance of machine learning as an internal measurement, and thus it is unnecessary to use a separate test data set for validation.

There are two parameters: the number of variables (M) in the random subset at each node and the number of trees (T) in the forest. The selection of parameter M has considerable influence on the final error rate. Both correlation between trees and the strength (classification accuracy) of individual trees in the forest increase with increase in M. The error rate is proportional to the correlation, but inversely proportional to the strength (Joelsson, Benediktsson, and Sveinsson 2008). M is generally set to the square root of the number of features (Gislason, Benediktsson, and Sveinsson 2006). Because random forest is fast and does not overfit, T can be as large as possible, although due to the memory limit of the machine, T is usually several hundred (Horning 2010); in this study, it was set to 100. According to Rodríguez-Galiano et al. (2011), random forest is insensitive to the value of M once the error has converged, and M alters classification accuracy only slightly.

Besides classification, random forest also provides a measure of variable importance. Variable importance measures the importance of the predictive variables (features) based on the permutation importance measure. Specifically, the process for an individual tree is defined as follows. (1) The original training data are re-sampled randomly to create

'bootstrap' samples ($Train_O$), each divided into an *inbag* and an OOB sample, from which a CART is constructed and validated. (2) Assuming that there are m features collected from remotely sensed data, a subset $M = \sqrt{m}$ features is randomly selected and tested for the best split at each node of CART based on the *Gini* impurity. The *Gini* impurity attains its maximum if each class occurs with equal probability; whereas it achieves its minimum if all randomly selected pixels at a node belong to only one class, that is, it is a pure node with a zero misclassification rate. Thus, CART selects the best split of the features as that split for which the reduction in impurity is highest. (3) To estimate a variable importance, a major voting scheme is used to evaluate votes from each tree in the forest. For example, for each tree j the OOB samples (OOB_j^f) of feature f are first run through to assess the classification accuracy and obtain the OOB error estimate (EB_j^f) on the OOB_j^f samples. Afterwards, a random permutation of feature f in the OOB_j^f samples is performed to obtain perturbed samples denoted by $(\mathrm{OOB}_j^f)'$ and the corresponding OOB error estimate$(\mathrm{EB}_j^f)'$. The importance of variable f per tree j is computed by

$$\mathrm{FI}_j^f = (\mathrm{OOB}_j^f)' - \mathrm{OOB}_j^f. \tag{1}$$

The importance score for feature f is then calculated as the mean importance over all trees:

$$\mathrm{FI}^f = \frac{1}{T}\sum_T \mathrm{FI}_j^f. \tag{2}$$

As a reliable indicator for providing measures of variable importance, the random variable permutation can simulate the absence of variable f from the forest (Guo et al. 2011) and provide varying different prediction accuracies before and after permutating the variable f. Following variable permutation, the importance score is represented by the decrease in correct class votes averaged over all trees. Thus, the greater the decrease in average predictive accuracy, the more important the feature f.

3.4. Feature selection based on random forests

Features from segments in the training samples are submitted to a process of selection. In this case, random forest-based feature selection was applied. To select optimal features, we leveraged a random forest algorithm (Breiman 2001) implemented in the 'Random Forest' package within the R environment.

Aiming to select optimal and uncorrelated features for the achievement of a good predictive performance, we employed a backward feature elimination proposed by Diaz-Uriarte and Alvarez de Andres (2006) for biological application by iteratively fitting a random forest. We first computed measures of feature importance to obtain an initial variable ranking and then proceeded with an iterative backward elimination of the least important variables. In each iteration, the least important features (by default, 20%) were eliminated and a new random forest was built by training with the remaining features for the assessment of OOB errors. The iterative procedure proceeded until the set of features leading to the lowest OOB errors of a forest was selected. The backward stepwise selection allows us to put all features into the random forest and eliminate unnecessary or partially correlated features one by one, according to the scores obtained from internal

validation of the random forest. Therefore, there is no prior knowledge of feature selection required.

3.5. Accuracy assessment

Both training and reference data are indispensable in supervised classification schemes. Training samples, as inputs to yielding a classification map, represent the spectral and geometrical signatures of each class. As a result, the choice of training method is as important as the choice of classifier, and it influences classification accuracy (Campbell 2006). Reference data are required for systematic comparison with the classification results for further accuracy assessment, referred to as external validation. To collect the ground reference data for this research, pixels were randomly selected by the 'Create Random Points' tool in ArcGIS 10 (ESRI®, Redlands, CA, USA), with a minimum allowable distance of 1 m. Each location was then investigated and labelled using higher-spatial-resolution remotely sensed data based on human visual interpretation. To reduce the effects of geometric misregistration, eight 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).

In order to determine the accuracy of classification results, an error matrix (Congalton 1981), also known as a confusion matrix, was built for each map. Producer's and user's accuracies for each class were calculated along with overall accuracy and kappa statistics (Congalton and Green 2009).

4. Results and discussion

4.1. Variable importance by global and per-class context

In this section, we describe experiments carried out for two study areas to calculate the importance of the contribution of each variable to both the general classification model and the classification of each category. Figure 2 shows the variable importance for training samples for each feature when all features were put into the random forest. The variable importance is demonstrated by mean OOB decrease in accuracy. According to the OOB mean decrease in accuracy in both cases, the most relevant among those 48 features include *nDSM*, eigenvalue-based anisotropy, and intensity GLCM. The contribution of certain aerial image-based GLCM measures (e.g. *Ent*., *Corr*., *S-M*) and lidar-based eigenvalue measures (e.g. planarity, linearity) is smaller than that of the rest of the input variables. Regarding the mean OOB decrease in accuracy, Figure 2 shows that the contribution of feature *nDSM* is the highest for all classes, with values up to 0.1344 and 0.1236 for the Mannheim and Niagara Falls data sets, respectively. In terms of decreased accuracy of OOB, height-related features such as eigenvalue-based anisotropy, N_{H}, NH_{FL}, are also very important.

Table 2 shows the importance of the contribution of each variable to the random forest classification, and demonstrates how those features impact the class separability of the classification scheme. For example, *nDSM* is the most important variable as it enables us to divide an image into high-rise (buildings and trees) and low-rise (ground and grass) objects. Thus, classification confusion between spectrally identical and elevationally different objects can be reduced, and classification accuracy significantly improved. As Mannheim and Niagara Falls are typical of the style of Western cities, there is a similar pattern of variable importance of the random forest. Moreover, variable lidar-NDVI has a special role in the classification of trees and grass. The reason behind this phenomenon is

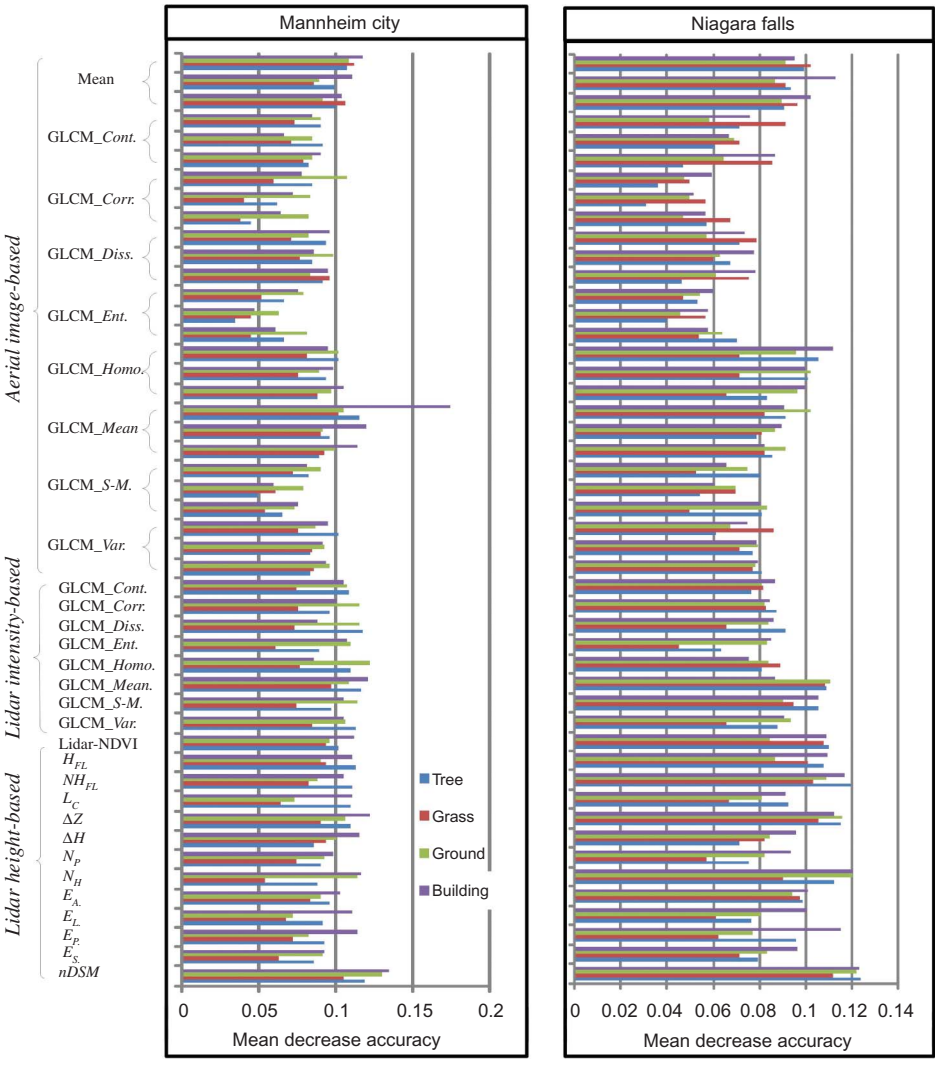


Figure 2. Variable importance of each feature demonstrated by mean OOB decrease in accuracy for two data sets, (*a*) Mannheim and (*b*) Niagara Falls.

that trees differ in both intensity and spectral content between the two cases because of seasonal differences. Like NDVI, lidar-NDVI generally helps to differentiate between urban areas and bare soils, which are covered by some grass in the spring season (Yuan and Bauer 2005).

To eliminate less important and better correlated variables/features, an iterative backward elimination scheme was used. In the present study, the number of trees (T) was set at 150 and the number of split variables at 4. Generally, a default value of split variables is a good choice regarding OOB error rate. Figure 3 shows the recursive feature elimination process for Mannheim and Niagara Falls using mean OOB decrease in accuracy. Mean OOB decrease accuracy for both study areas rises with a decrease in numbers of variables from 48 to 10; the kappa statistics of the object-based classification increase accordingly,

Table 2. Per-class variable importance in terms of mean OOB decrease in accuracy, ranking of importance from top to bottom.

Mannheim				Niagara Falls			
Tree	Grass	Ground	Building	Tree	Grass	Ground	Building
Intensity-GLCM-*Mean*	*nDSM*	*nDSM*	*nDSM*	*nDSM*	*nDSM*	*nDSM*	*nDSM*
nDSM	Intensity-GLCM-*Mean*	N_H	ΔZ	NH_{FL}	Intensity-GLCM-*Mean*	N_H	N_H
NH_{FL}	NH_{FL}	ΔZ	Intensity-GLCM-*Mean*	ΔZ	Lidar-NDVI	ΔZ	NH_{FL}
N_H	ΔZ	Intensity-GLCM-SM	E_A	N_H	ΔZ	Intensity-GLCM-*Mean*	E_P
ΔZ	Lidar-NDVI	NH_{FL}	Intensity-GLCM-SM	Lidar-NDVI	NH_{FL}	NH_{FL}	ΔZ
Lidar-NDVI	H_{FL}	E_A	NH_{FL}	Intensity-GLCM-*Mean*	H_{FL}	RGB-GLCM-*Homo*	H_{FL}
E_A	RGB-GLCM-*Homo*	Intensity-GLCM-*Mean*	N_H	H_{FL}	E_A	E_A	Lidar-NDVI
Intensity-GLCM-*S-M*	E_A	H_{FL}	Lidar-NDVI	Intensity-GLCM-*S-M*	Intensity-GLCM-*S-M*	RGB-GLCM-*Mean*	RGB-GLCM-*Homo*
RGB-GLCM-*Homo*	Intensity-GLCM-*S-M*	RGB-GLCM-*Homo*	H_{FL}	RGB-GLCM-*Homo*	N_H	Intensity-GLCM-*S-M*	Intensity-GLCM-*S-M*
H_{FL}	N_H	Lidar-NDVI	RGB-GLCM-*Homo*	E_A	RGB-GLCM-Cont	H_{FL}	E_A

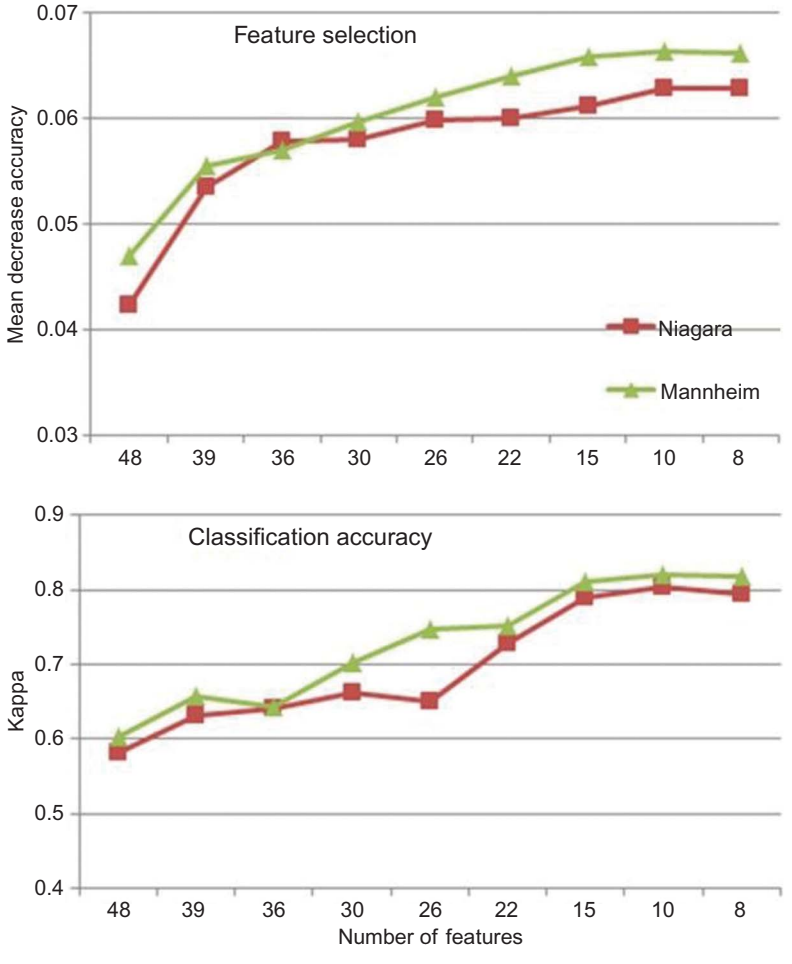


Figure 3. Classification accuracy of two data sets at different number of features selected by iterative feature elimination.

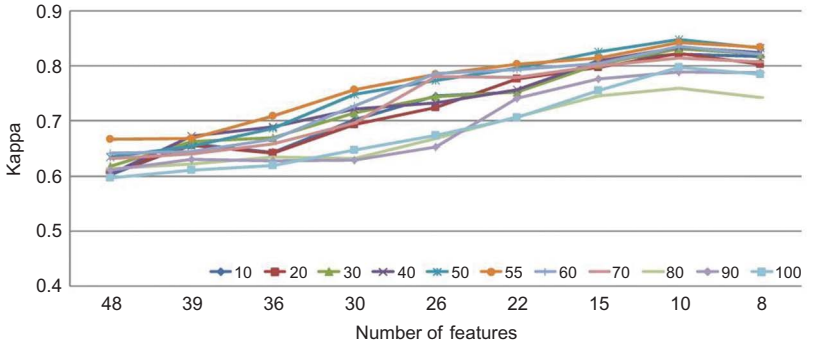
because the backward feature selection would eliminate some redundant and partially correlated features. However, the kappa statistics slowly decrease when the number of features is below 10, indicating that an excessive elimination of features can be counterproductive. For the two cases, the 10 potent features retained included lidar-NDVI, lidar height-based measures (*nDSM*, ΔZ, NH_{FL}, ΔH), lidar intensity GLCM (*Mean* and *S-M*), and aerial image-based GLCM (*Homo*.). Correspondingly, classification accuracies of each feature group at the fine scale of 10 are shown at the foot of Figure 3. Although the features selected change from one city to the other, classification accuracies exhibit similar trends. Figure 3 shows that the selected minimal features with a low OOB error rate can improve classification performance.

4.2. Effects of scale on feature importance and selection

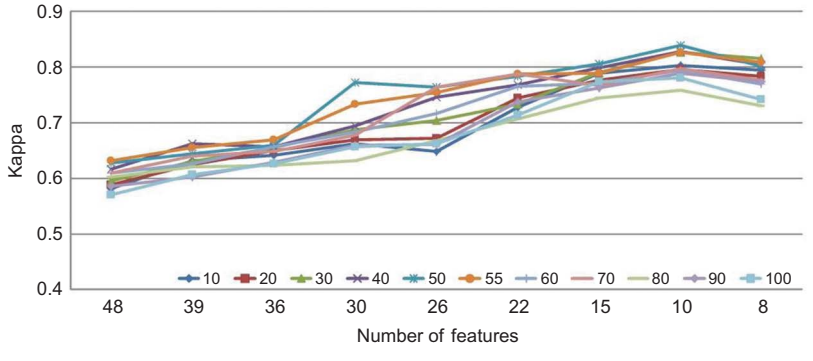
The quality of segmentation largely determines the accuracy of final urban classification results because a segment rather than a pixel is considered as the unit to be investigated. As increasing segmentation scales lead to larger segment sizes, the features selected based on the random forest will be different between scales and, as a result, the classification

results will be affected. The larger the segments, the fewer the samples selected for the estimation of the OOB errors. According to Diaz-Uriarte and Alvarez de Andres (2006), standard OOB errors increase with decrease in the number of samples because a feature selection method can deteriorate strongly if the number of features is larger than the number of samples. Although the number of features used in the present study is 48, the number of samples was at least 2–3-fold the number of features, even at the largest segmentation scale.

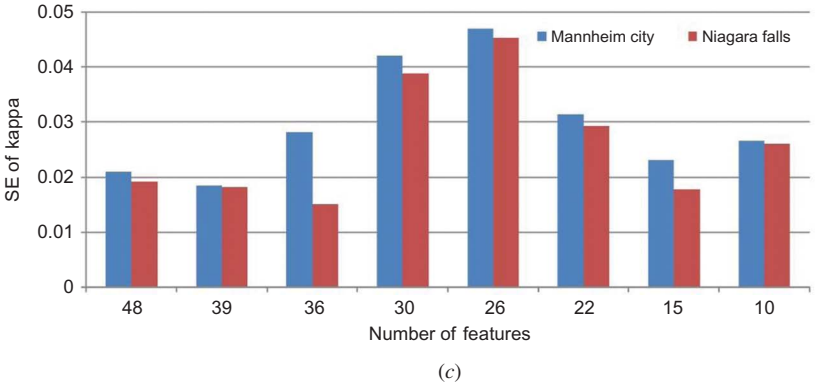
Figure 4 shows the classification accuracy of the two cases at eleven segmentation scales when features are gradually eliminated in terms of OOB errors. Theoretically, the



(*a*)



(*b*)



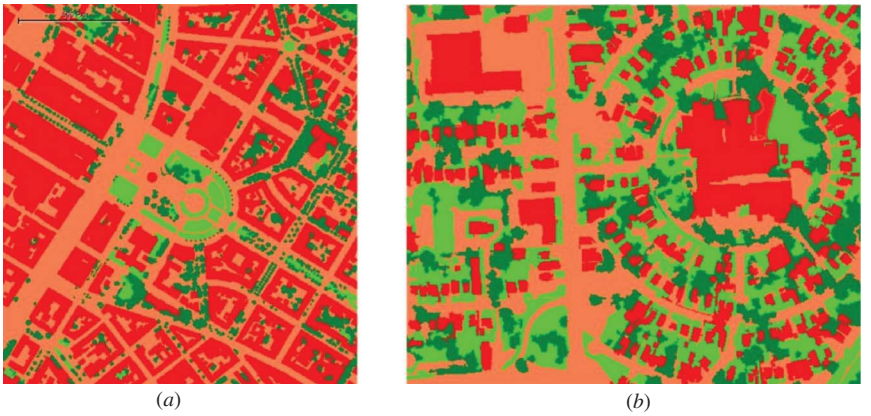
(*c*)

Figure 4. Kappa statistics at different segmentation scales for (*a*) Mannheim data set, (*b*) Niagara Falls data set, and (*c*) standard error for eleven scales.

predictive accuracy could be increased by using a smaller number of features with fewer correlated relationships. As shown in Figures 4(*a*) and (*b*), the classification accuracies of the two cities exhibit the same upward trend when the number of selected features is between 48 and 10. This is because the method of iterative feature elimination reduces some redundant and partially correlated features. However, the classification accuracy slowly decreases when the number of selected features is below 10, indicating that an excessive elimination of features can be counterproductive. In fact, at a certain segmentation scale, classification accuracy would reach its peak at a set number of features. For example, for most segmentation scales, the best performance would be obtained when the number of features is 10.

The classification accuracy of these two cases decreases at larger segmentation scales for the same number of selected features. However, the standard error (SE) of the kappa statistics for all eleven scales changed slightly, from 0.015 to 0.047 (Figure 4(*c*)). Although the features selected depend on the study scene and data characteristics such as segmentation scale, some features common to the two cases are significant for classification. Undoubtedly, geometric features such as *nDSM*, N_{H}, NH_{FL}, and lidar-NDVI have a considerably impact on all test cases. The variable importance ranking of these features displays low variability among different segmentation scales. The lidar-derived *nDSM* helps to reduce OOB errors because it contains elevated objects. Thus, the use of *nDSM* can effectively reduce classification confusion between spectrally identical or similar objects. Furthermore, both N_{H} and NH_{FL} (a lidar-based vegetation index with respect to the first and last echoes) are ranked higher in most cases, indicating that lidar-derived geometric features significantly contribute to classification results. Lidar-NDVI, another indicator for vegetation detection that combines lidar intensity and spectral information, also plays a critical role in the classification process. In addition, the most suitable texture measures – image-based GLCM-*Homo* and lidar intensity GLCM-*Mean* – are included in the features selected for recognizing man-made objects and grass, respectively.

Figure 5 shows the resultant image of the object-based classification. Visual inspection indicates that the four classes of interest are separated from each other very well. Buildings are completely detected and classified. Other objects (e.g. cars, poles) not belonging to the four classes of interest are removed as miscellanies. However, some classification errors of trees occur along streets because these trees are not high enough to contribute to the separation of tree objects in the MRIS segmentation process.



(*a*) (*b*)

Figure 5. Object-based classification maps of (*a*) Mannheim data set and (*b*) Niagara Falls data set.

4.3. Comparison of pixel-based and object-based classification

To further assess the object-based classification described above, an experiment was conducted to compare the performance of the proposed object-based classification method with the pixel-based classification method, both of which were based on the random forest. To obtain the desired classification accuracy, the number of training pixels should at least be equal to tenfold the number of variables used in the classification model (Jensen 2005), and the number of pixels in such a group should not be larger than 10 (Congalton, Oderwald, and Mead 1983) for a parametric classification approach. The final accuracy assessment of the object-based classification was performed at a segmentation scale (55) selected. According to previous analyses, the number of features selected was 10 for acceptable classification performance. Similarly, we also employed the method of iterative feature elimination to obtain the optimal features for pixel-based classification. The experiments showed that the optimal number of feature was 12 at the highest mean decrease in accuracy.

Table 3 compares pixel-based and object-based classification. The results show that object-based classification attains higher overall accuracy, kappa coefficient, producer's accuracy, and user's accuracy than the pixel-based classification for most classes. For the Mannheim and Niagara Falls data sets, the overall accuracy of the object-based classification was 4.5% and 3.9% higher than that of the pixel-based classification, respectively. The kappa coefficients of pixel-based classification were 0.3 and 0.5 higher than those of object-based classification for these two data sets.

In object-based classification, trees had producer's accuracy of 92.1% and 89.6%, and user's accuracy of 78.7% and 80.2% for the Mannheim and Niagara Falls data sets, respectively. This shows an increase of 3.0% and 8.8% in producer's accuracies and 1.7% and 5.1% in user's accuracies, respectively. This improvement is probably due to the enhanced quality of image segmentation that overcomes dramatic spectral variation of trees in the scenes by integrating lidar elevation and spectral information.

In regard to grass, compared with pixel-based classification, object-based classification showed an increase of 0.2% and 3.4% in producer's accuracy, and a growth of 2.0% and 3.9% in user's accuracy for the Mannheim and Niagara Falls data sets, respectively. Although we used variable lidar-NDVI, a primary indicator to identify grass, neither classification methods could accurately distinguish grass from low-rise vegetation in user's accuracy.

Building achieved low classification accuracy in pixel-based classification, probably due to the proximity of buildings to trees and confusion between buildings and bare ground. This confusion is not too surprising, as these classes have a certain spectral similarity. In object-based classification, Building showed an increase of 0.9% and 1.4% in producer's accuracy and increase of 0.3% and 14.5% in user's accuracy for the Mannheim and Niagara Falls data sets, respectively. This accuracy improvement is due to the integration of elevation data and spectral information and the implementation of variable *nDSM*, both of which could reduce classification confusion between buildings and bare ground.

Object-based classification outperformed pixel-based classification except for the user's accuracy of bare ground in the Mannheim data set, where it achieved 3% lower accuracy. Misclassification errors might have been caused by some grassy areas mistakenly segmented and classified as bare ground (Figure 5(*a*)), despite the use of variable lidar-NDVI.

The experimental results indicate that the object-based method using random forest-based feature selection can produce an incremental improvement in classification accuracy when applied to lidar data integrated with high-spatial-resolution aerial imagery. However,

Table 3. A comparison of accuracy assessment between pixel-based and object-based classification.

				Producer's accuracy (%)				User's accuracy (%)			
Test sites	Method	Overall accuracy (%)	Kappa	Building	Tree	Ground	Grass	Building	Tree	Ground	Grass
Mannheim	Pixel-based	81.6	0.81	85.6	89.1	83.3	87.7	84.3	76.6	96.5	60.8
	Object-based	86.1	0.83	86.5	92.1	85.5	87.9	84.6	78.7	93.5	62.8
Niagara Falls	Pixel-based	82.4	0.79	86.1	80.8	75.4	85.3	70.2	75.1	80.3	59.3
	Object-based	86.3	0.84	87.5	89.6	85.1	88.7	84.7	80.2	94.4	63.2

due to landscape complexity, classification performance could be weakened by segmentation. This challenge can be overcome by the further development of a robust segmentation algorithm that regenerates segmentation results using the best feature subset retrieved by random forest-based feature selection.

5. Conclusions

Previous classification schemes from lidar data and aerial imagery are highly reliant on a subjective selection of suitable features, making it difficult to adapt them to new city types and data sets. To resolve such issues, this study employed random forest and image segmentation for feature selection and land-cover classification. A variety of features from both lidar data and imagery were investigated, and the optimal features selected by the random forest were calculated through the lowest mean OOB decrease in accuracy. Although the optimal features selected varied from case to case, a number of lidar, spectral and intensity-based features were selected in both cases in this study. Lidar-derived *nDSM*, NH_{FL} (a variant of vegetation index calculated by first and last echoes), lidar-NDVI, *intensity-based* GLCM-*Mean*, and spectral-based GLCM measures provide complementary information to identify the classes of interest, which are useful for urban mapping using lidar data and aerial imagery, even for other types of remotely sensed data.

By grouping neighbouring pixels of the same characteristics into segments, object-based classification can overcome the salt-and-pepper effect that often occurs in pixel-based classification. This article applied MRIS to obtain a series of segments at different scales. The impact of segmentation scales on random forest-based feature selection was investigated. Although larger scales produced smaller training samples for the random forest, the optimal features selected at eleven scales had little influence on classification accuracies. Moreover, by integrating the lidar-derived range image into the segmentation stage as an additional band, the quality of image segmentation was improved. Comparison between pixel-based classification and the object-based classification demonstrates that object-based classification was superior by 0.02 and 0.05 in kappa statistics, and by 3.9% and 4.5% in overall accuracy for the Mannheim and Niagara Falls data sets, respectively. Therefore, improvement in urban land-cover classification accuracy can be improved not only by the combination of lidar data and aerial imagery, but also by the integration of random forest-based feature selection and object-based classification.

Acknowledgements

This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant awarded to Professor Jonathan Li, and National Natural Science Foundation of China (NSFC) grants (No. 41071257; No. 41001305). The authors would like to thank the anonymous reviewers for their constructive comments and suggestions, which have improved the article significantly.

References

Alonso, M. C., and J. A. Malpica. 2010. “Satellite Imagery Classification with Lidar Data.” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Kyoto, 38: 730–735.

Arefi, H., M. Hahn, and J. Lindenberger. 2003. “Lidar Data Classification with Remote Sensing Tools.” In *Proceedings of the ISPRS Commission IV Joint Workshop: Challenges in Geospatial Analysis, Integration and Visualization II*, Stuttgart, September 8–9, 131–136.

Benedict, M. A., and E. T. Mcmahon. 2002. “Green Infrastructure: Smart Conservation for the 21st Century.” *Renewable Resources Journal* 20: 12–17.

Benz, U. C., P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen. 2004. “Multiresolution Object-Oriented Fuzzy Analysis of Remote Sensing Data for GIS-Ready Information.” *ISPRS Journal of Photogrammetry and Remote Sensing* 58: 239–258.

Blaschke, T. 2010. “Object Based Image Analysis for Remote Sensing.” *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 2–16.

Brandtberg, T. 2007. “Classifying Individual Tree Species Under Leaf-Off and Leaf-on Conditions Using Airborne Lidar.” *ISPRS Journal of Photogrammetry and Remote Sensing* 61: 325–340.

Breiman, L. 1996. “Bagging Predictors.” *Machine Learning* 24: 123–140.

Breiman, L. 2001. “Random Forests.” *Machine Learning* 45: 5–32.

Brenner, C. 2010. “Building Extraction.” In *Airborne and Terrestrial Laser Scanning*, edited by G. Vosselman, and H.-G. Maas, 169–211. Dunbeath: Whittles.

Campbell, J. B. 2006. *Introduction to Remote Sensing*, 343p. New York: The Guilford Press.

Chehata, N., L. Guo, and C. Mallet. 2009. “Airborne Lidar Feature Selection for Urban Classification Using Random Forests.” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science* 38 (3W8): 207–212.

Chen, Y., W. Su, J. Li, and Z. Sun. 2009. “Hierarchical Object Oriented Classification Using Very High Resolution Imagery and LiDAR Data Over Urban Areas.” *Advances in Space Research* 43: 1101–1110.

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

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

Congalton, R. G., R. G. Oderwald, and R. A. Mead. 1983. “Assessing Landsat Classification Accuracy Using Discrete Multivariate Statistical Techniques.” *Photogrammetric Engineering & Remote Sensing* 49: 1671–1678.

Diaz-Uriarte, R., and S. Alvarez de Andres. 2006. “Gene Selection and Classification of Microarray Data Using Random Forest.” *BMC Bioinformatics* 7. doi:10.1186/1471-2105-7-3.

Duveiller, G., P. Defourny, B. Desclée, and P. Mayaux. 2008. “Deforestation in Central Africa: Estimates at Regional, National and Landscape Levels by Advanced Processing of Systematically-Distributed Landsat Extracts.” *Remote Sensing of Environment* 112: 1969–1981.

Gamanya, R., P. De Maeyer, and M. De Dapper. 2009. “Object-Oriented Change Detection for the City of Harare Zimbabwe.” *Expert Systems with Applications* 36: 571–588.

Germaine, K. A., and M. Huang. 2011. “Delineation of Impervious Surface from Multispectral Imagery and LiDAR Incorporating Knowledge Based Expert System Rules.” *Photgrammetric Engineering & Remote Sensing* 77: 77–87.

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson. 2006. “Random Forests for Land Cover Classification.” *Pattern Recognition Letters* 27: 294–300.

Guan, H., Z. Ji, L. Zhong, J. Li, and Q. Ren. 2013. “Partially Supervised Hierarchical Classification for Urban Features from Lidar Data with Aerial Imagery.” *International Journal of Remote Sensing* 34 (1): 190–210.

Guan, H., J. Li, and M. A. Chapman. 2011. “Urban Thematic Mapping by Integrating Lidar Point Cloud with Colour Imagery.” *GEOMATICA* 65: 375–385.

Guan, H., J. Li, M. A. Chapman, L. Zhong, and Q. Ren. 2011. “Support Vector Machine for Urban Land-use Classification Using Lidar Point Clouds and Aerial Imagery.” In *Lidar and Radar Mapping*, Nanjing, May 26–28, 8p.

Guo, L., N. Chehata, C. Mallet, and S. Boukir. 2011. “Relevance of Airborne Lidar and Multispectral Image Data for Urban Scene Classification Using Random Forests.” *ISPRS Journal of Photogrammetry and Remote Sensing* 66: 56–66.

Haala, N., and V. Walter. 1999. “Classification of Urban Environments Using LiDAR and Color Aerial Imagery.” *International Archives of Photogrammetry and Remote Sensing* 32 (7-4-3 W6), Valladolid, June 3–4, 76–82.

Haralick, R. M., K. Shanmugam, and I. Dinstein. 1973. “Textural Features for Image Classification.” *IEEE Transactions on Systems, Man and Cybernetics* 3: 610–621.

Horning, N. 2010. “Random Forests: An Algorithm for Image Classification and Generation of Continuous Fields Data Sets.” In *Proceeding of International Conference on Geoinformatics*

for Spatial Infrastructure Development in Earth and Allied Sciences, Hanoi, December 9–11, Accessed September 15. http://wgrass.media.osaka-cu.ac.jp/gisideas10/viewpaper.php?id=342.

Huang, M., S. Shyue, L. Lee, and C. Kao. 2008. "A Knowledge-Based Approach to Urban Feature Classification Using Aerial Imagery with Lidar Data." *Photogrammetric Engineering & Remote Sensing* 74: 1473–1485.

Jensen, J. R. 2005. *Introductory Digital Image Processing*. 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Joelsson, S. R., J. A. Benediktsson, and R. Sveinsson. 2008. "Random Forest Classification of Remote Sensing Data." In *Image Processing for Remote Sensing*, edited by C. H. Chen, 61–78. Boca Raton, FL: CRC Press.

Julesz, B. 1962. "Visual Pattern Discrinmination." *IRE Transactions on Information Theory* 8: 84–92.

Jutzi, B., and H. Gross. 2009. "Normalization of LiDAR Intensity Data Based on Range and Surface Incidence Angle." In *International Archives of Photogrammetry and Remote Sensing 38(Part3/W8)*, Paris, September 1–2, 213–218.

Ke, Y., L. J. Quackenbush, and J. Im. 2010. "Synergistic Use of QuickBird Multispectral Imagery and LIDAR Data for Object-Based Forest Species Classification." *Remote Sensing of Environment* 114: 1141–1154.

Khoshelham, K., C. Nardinocchi, E. Frontoni, A. Mancini, and P. Zingaretti. 2010. "Performance Evaluation of Automated Approaches to Building Detection in Multi-Source Aerial Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 123–133.

Lehrbass, B., and J. Wang. 2012. "Urban Tree Cover Mapping with Relief-Corrected Aerial Imagery and Lidar." *Photogrammetric Engineering & Remote Sensing* 78: 473–484.

Li, J., and H. Guan. 2011. "3d Building Reconstruction from Lidar Point Clouds Fused with Aerial Imagery." In *Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment*, edited by X. Yang, 75–92. Oxford: Wiley-Blackwell.

Maas, H., and G. Vosselman. 1999. "Two Algorithms for Extracting Building Model from Raw Laser Altimetry Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 54: 153–163.

Mallet, C., F. Bretar, and U. Soergel. 2008. "Analysis of Full-Waveform Lidar Data for Classification of Urban Areas." *Photogrammetrie Fernerkundung Geoinformation* 5: 337–349.

Melzer, T. 2007. "Non-Parametric Segmentation of ALS Point Clouds Using Mean Shift." *Journal of Applied Geodesy* 1: 159–170.

Pineda-Bautista, B. B., J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad. 2011. "General Framework for Class-Specific Feature Selection." *Expert System Application* 38: 10018–10024.

Rodríguez-Galiano, V. F., M. Chica-olmo, F. Abarca-Hernandez, P. M. Atkinson, and C. Jeganathan. 2012. "Random Forest Classification of Mediterranean Land Cover Using Multi-Seasonal Imagery and Multi-Seasonal Texture." *Remote Sensing of Environment* 121: 93–107.

Rottensteiner, F., J. Trinder, S. Clode, and K. Kubik. 2005. "Using the Dempster-Shafer Method for the Fusion of LIDAR Data and Multi-Spectral Images for Building Detection." *Information Fusion* 6: 283–300.

Secord, J., and A. Zakhor. 2007. "Tree Detection in Urban Region Using Aerial Lidar and Image Data." *IEEE Geoscience and Remote Sensing Letters* 4: 196–200.

Shan, J., and S. Lee. 2003. "Combining LiDAR Elevation Data and Ikonos Multispectral Imagery for Coastal Classification Mapping." *Marine Geodesy* 26: 117–127.

Stumpf, A., and N. Kerle. 2011. "Object-Oriented Mapping of Landsides Using Random Forests." *Remote Sensing of Environment* 115: 2564–2577.

Vögtle, T., and E. Steinle. 2003. "On the Quality of Object Classification and Automated Building Modeling Based on Laser Scanning Data." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Dresden, 34 (3W13): 149–155.

Walter, V. 2004. "Object-Based Classification of Remote Sensing Data for Change Detection." *ISPRS Journal of Photogrammetry and Remote Sensing* 58: 225–238.

Wang, X., B. Waske, and J. A. Benediktsson. 2009. "Ensemble Methods for Spectral-Spatial Classification of Urban Hyperspectral Data." *Proceedings of IGARSS, Cape Town, South Africa* 4: 944–947.

Yu, H., G. Cheng, X. Ge, and X. Lu. 2011. "Object Oriented Land Cover Classification Using ALS and GeoEYE Imagery Over Mining Area." *Transactions of Nonferrous Metals Society of China* 21: s733–s737.

Yuan, F., and M. E. Bauer. 2005. "Comparison of Impervious Surface Area and Normalized Difference Vegetation Index as Indicators of Surface Urban Heat Island Effects in Landsat Imagery." *Remote Sensing of Environment* 106: 375–386.

Zhou, W., and A. Troy. 2008. "An Object-Oriented Approach for Analysing and Characterizing Urban Landscape at the Parcel Level." *International Journal of Remote Sensing* 29: 3119–3135.

Zimble, D. A., D. L. Evans, G. C. Carlson, R. C. Parker, S. C. Grado, and P. D. Gerard. 2003. "Characterizing Vertical Forest Structure Using Small-Footprint Airborne LiDAR." *Remote Sensing of Environment* 87: 171–182.