ISPRS Journal of Photogrammetry and Remote Sensing 66 (2011) 56–66

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests

Li Guo[a], Nesrine Chehata[a,b,*], Clément Mallet[b], Samia Boukir[a]

[a] Institut EGID, Université de Bordeaux, Laboratoire GHYMAC 1 allée F. Daguin 33670 Pessac, France

[b] Université Paris Est, IGN, Laboratoire MATIS 73 avenue de Paris 94165 Saint-Mandé, France

ARTICLE INFO

Article history:
Received 8 April 2010
Received in revised form
19 August 2010
Accepted 20 August 2010
Available online 22 September 2010

Keywords:
Lidar
Multispectral image
Urban
Random forests
Variable importance

ABSTRACT

Airborne lidar systems have become a source for the acquisition of elevation data. They provide georeferenced, irregularly distributed 3D point clouds of high altimetric accuracy. Moreover, these systems can provide for a single laser pulse, multiple returns or echoes, which correspond to different illuminated objects. In addition to multi-echo laser scanners, full-waveform systems are able to record 1D signals representing a train of echoes caused by reflections at different targets. These systems provide more information about the structure and the physical characteristics of the targets. Many approaches have been developed, for urban mapping, based on aerial lidar solely or combined with multispectral image data. However, they have not assessed the importance of input features. In this paper, we focus on a multi-source framework using aerial lidar (multi-echo and full waveform) and aerial multispectral image data. We aim to study the feature relevance for dense urban scenes. The Random Forests algorithm is chosen as a classifier: it runs efficiently on large datasets, and provides measures of feature importance for each class. The margin theory is used as a confidence measure of the classifier, and to confirm the relevance of input features for urban classification. The quantitative results confirm the importance of the joint use of optical multispectral and lidar data. Moreover, the relevance of full-waveform lidar features is demonstrated for building and vegetation area discrimination.

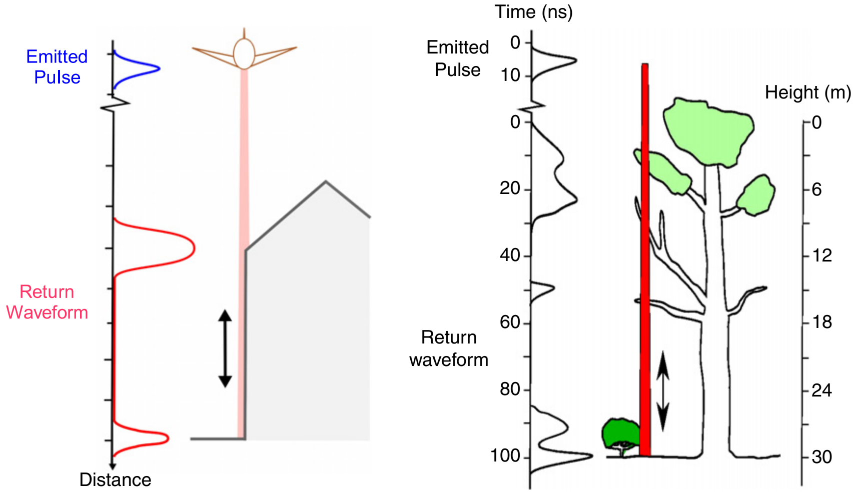
1. Introduction

Airborne lidar systems have become a source for the acquisition of altimeter data. The development of various approaches based on lidar data for urban mapping has been an important issue for the last few years (Brenner, 2010). Many authors have shown the potential of multi-echo lidar data for urban area analysis and building extraction based on filtering and segmentation processes (Matikainen et al., 2003; Sithole and Vosselman, 2004; Rottensteiner et al., 2007; Matei et al., 2008). Classification is used for urban mapping to focus on building or vegetation areas, before the modeling step (Haala and Brenner, 1999; Poullis and Yu, 2009; Zhou and Neumann, 2009). Several classification methods were applied to lidar data for urban scenes such as the unsupervised Mean-Shift algorithm (Melzer, 2007), supervised classification such as Support Vector Machines (SVM) (Secord and Zakhor, 2007; Mallet et al., 2008), and a cascade of binary classifiers based on 3D shape analysis (Carlberg et al., 2009). Lidar classification can be based on geometric and textural features (Matikainen et al., 2003). Other methods include the lidar intensity (Charaniya et al., 2004), and combined lidar and multispectral data (Rottensteiner et al., 2007; Secord and Zakhor, 2007).

Recently, the full-waveform (FW) lidar technology (Mallet and Bretar, 2009) has emerged with the ability to record 1D signals representing multiple modes (echoes) caused by reflections at different targets (*cf.* Fig. 1). Thus, in addition to range measurements, further physical properties of targets may be revealed by waveform processing and echo fitting. In urban scenes, the potential of such data has been essentially investigated for urban vegetation with high density point clouds. In Gross et al. (2007) and Wagner et al. (2008), geometric and FW lidar features are derived from a FW 3D point cloud and used jointly to discriminate vegetated areas. In Rutzinger et al. (2008), the authors present an object-based analysis of FW lidar point cloud to extract urban vegetation. The 3D point cloud is first slightly over-segmented using a seeded region growing algorithm based on the echo width. Each segment is characterized by basic statistics (minimum, maximum, standard deviation, etc.) computed for the selected point features: amplitude, echo width and geometrical attributes. A supervised classification per statistical tree decision is then applied. In Höfle and Hollaus (2010), an improved echo ratio feature is computed

[*] Corresponding author at: Institut EGID, Université de Bordeaux, Laboratoire GHYMAC 1 allée F. Daguin 33670 Pessac, France. Tel.: +33 662007836.
E-mail addresses: nesrine.chehata@egid.u-bordeaux3.fr, nesrine.chehata@ign.fr (N. Chehata).

0924-2716/$ – see front matter © 2010 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2010.08.007



(a) Building area.

(b) Vegetation area.

Fig. 1. Transmitted and received signals with a small-footprint FW lidar.

allowing a better vegetation discrimination. A rule-based classification is refined using FW echo amplitude and width to discriminate non-vegetation objects such as building walls, roof edges, and powerlines. Mallet et al. (2008) studied the contribution of FW lidar data for urban scene classification using 3D features and SVM classification.

All these works focus on the classification process but do not analyze the relevance of input features. The contribution of lidar data in comparison with multispectral images is not quantified, nor is the importance of FW lidar features for urban scenes. Our objective is to study the relevance of multi-source data composed of lidar features (multi-echo (ME) and full waveform (FW)) and multispectral *RGB* features for mapping urban scenes. Four urban classes are considered: buildings, vegetation, artificial ground, and natural ground. Artificial ground gathers all kinds of streets and street items such as cars and traffic lights whereas natural ground includes grass, sand and bare-earth regions. To achieve this goal, the Random Forests classifier is chosen. This algorithm is well suited to a multi-source framework and is able to process large datasets. Besides, it measures feature importance so that its contribution can be examined for the different classes we consider. In our previous work Chehata et al. (2009b), feature importance was studied for 2D ME and FW lidar features and multispectral data. In Chehata et al. (2009a), 21 lidar features were generated and analyzed, and the influence of the 2D window size was studied. In this work, we present a global framework to study the relevance of features for classification using the Random Forests classifier. Besides, to confirm these observations, the margin theory is used as a confidence measure of the classifier, which consequently helps in evaluating the suitability of input features. The methodology is applied in this case to lidar and multispectral data. The remainder of this paper is organized as follows. The lidar and optical features are introduced in Section 2. In Section 3, Random Forests classification, the feature importance measure, and a new margin are presented. Experimental results are then given and discussed in Section 5. Finally, conclusions and perspectives are drawn.

2. Multi-source airborne lidar and image features

The multi-source data is composed of an orthoimage and a full-waveform lidar dataset. The orthoimage is composed of three multispectral bands in the visible domain: Red, Green and Blue. Lidar and image data are complementary: the optical image provides high spatial resolution and multispectral reflectances in the visible domain, while lidar uses the infrared domain and provides 3D geometric information, but the data is under-sampled. Moreover, it has the ability to penetrate the vegetation, giving more information about these areas (Brenner, 2010). Finally, FW lidar provides more information about the physical properties of the targets. The data are georeferenced and processed in the same geographic projection system. In order to combine data from sources with distinct geometries, lidar points are projected onto a regular 2D raster map. One image is generated per feature used for classification. Red (R), Green (G) and Blue (B) channels of the orthoimage are used as three independent optical features.

The feature vector is composed of twelve features: three optical features R, G and B, five multi-echo lidar features and four full-waveform lidar features. Lidar features are obtained by lidar waveform modeling. It consists in decomposing the waveform into a sum of components or echoes, so as to characterize the different targets along the path of the laser beam. A parametric approach is chosen. Parameters of an analytical function are estimated for each detected peak in the signal. Generally, the received signal is decomposed by fitting Gaussians to the waveform (Wagner et al., 2006; Reitberger et al., 2009). The waveform fitting is processed by an iterative Levenberg–Marquardt technique. This, first, leads to multiple echo detection and range measurements. In our study, ME lidar features are derived from a FW lidar dataset providing a higher altimetric accuracy. However, they can be obtained directly using a multi-echo lidar system.

For each pixel, the lidar features are computed using the 3D points included in a given cylindrical neighborhood ν_P, centered at the current pixel P and defined by the parameter r (*cf.* Fig. 2(a)).

The raster cell spacing c and the cylinder radius r are chosen with respect to the following:

- the 3D point density: a minimal number of lidar points is necessary to compute an unbiased local plane Π_P;
- the contrast between objects that we aim to retrieve: a small value of c combined with a high value of r can lead, with a dense point cloud, to smoothed images, and to mix in a single pixel, different kinds of objects (see Fig. 2(a)). In such cases, the geometric spatial features may be affected, thus providing biased values.

We assume that a 3D neighborhood should include at least five points to process lidar features. The minimal radius is fixed in concordance with the 3D point density. The maximal radius equals

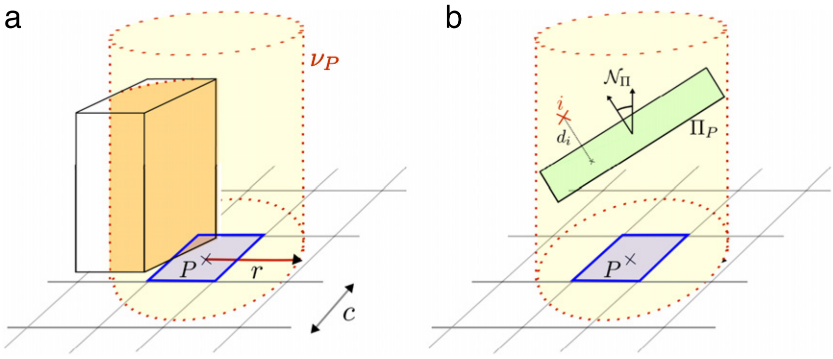


Fig. 2. 3D neighborhood for lidar feature computation.

the maximal object size in the image. For each 3D neighborhood, the radius is increased iteratively until validating our assumption. Values of c and r for our dataset are given in Section 4.

The five multi-echo lidar features are spatial. In urban scenes, most objects can be described by planar surfaces such as building roofs and roads. The planarity of the local neighborhood should help in discriminating buildings from vegetation. The local plane Π_P is estimated using a robust M-estimator with norm $L_{1.2}$ (Xu and Zhang, 1996) on the points included in ν_P. Such a norm provides a plane estimator scarcely affected by outlying data, such as superstructures for buildings, and low-rise objects for ground areas. Several features are derived from the computation of Π_P.

- Δz: height difference between the current lidar point P and the lowest point found in a large cylindrical volume whose radius r depends on the size of the largest structure in the area of interest. It has been experimentally set to 15 m. This feature will help in discriminating between ground and off-ground objects without a digital terrain model (DTM) estimation.
- N_z: deviation angle of the normal vector of the local plane Π_P from the vertical direction (see Fig. 2(b)). This feature highlights the ground with the lowest values.
- $\mathcal{R}_z$: residuals of the local plane estimated within the small radius cylinder. Residuals $\mathcal{R}_z$ are calculated with respect to the estimated plane as follows:

$$\mathcal{R}_z = \sum_{i\in\nu_P} \frac{(d_i)^l}{l} \qquad (1)$$

where d_i is the distance between the lidar point $P_i \in \nu_P$ (ν_P is the local 3D neighborhood) and the plane Π_P (see Fig. 2(b)). L_l is the norm used for estimating the 3D plane. Here $l = 1.2$. Residuals should be high for vegetation.

A single emitted laser pulse may generate several echoes from objects located at different positions inside the pulse conical 3D volume. In urban scenes, this is particularly interesting for vegetated areas and building edges since several echoes will be recorded per emitted pulse. Consequently, features based on the number of echoes should be relevant for urban scene classification. Two features have therefore been selected.

- N: total number of echoes within each waveform of the current pixel P. This feature will be prominent for vegetation and building facades (*cf.* Fig. 3(e)).
- N_e: normalized number of echoes obtained by dividing the echo number (position of the echo within the waveform) by the total number of echoes within the waveforms of P. This feature highlights the vegetation since multiple reflections can occur on it (*cf.* Fig. 3(f)).

The remaining features are more specific to FW lidar data. The received signal is generally decomposed by fitting Gaussians to the waveform. However, in urban areas, the characteristics of peaks may differ significantly due to multiple geometric and radiometric effects of the targets (*e.g.*, roof slopes and various materials). Consequently, other modeling functions have been proposed. Chauve et al. (2007) have improved the signal fitting using the generalized Gaussian function. We used the latter modeling to retrieve the following FW lidar features.

- A (echo amplitude): high amplitude values can be found on building roofs, on gravel, and on cars. Asphalt and tar streets have low values. The lowest values correspond to vegetation due to the penetration of the laser pulse and the signal attenuation (*cf.* Fig. 3(g)).
- w (echo width): higher values correspond to vegetation since it spreads the lidar pulses. A narrow pulse is likely to correspond to ground and buildings. However, the width value increases with the slope (*cf.* Fig. 3(h)).
- σ (echo cross-section (Wagner et al., 2008)): the values are high for buildings, medium for vegetation, and low for artificial ground (*cf.* Fig. 3(i)).
- α: echo shape, describing how locally flattened the echo is. Chauve et al. (2007) showed that very low and high shape values correspond respectively to building roofs and vegetation.

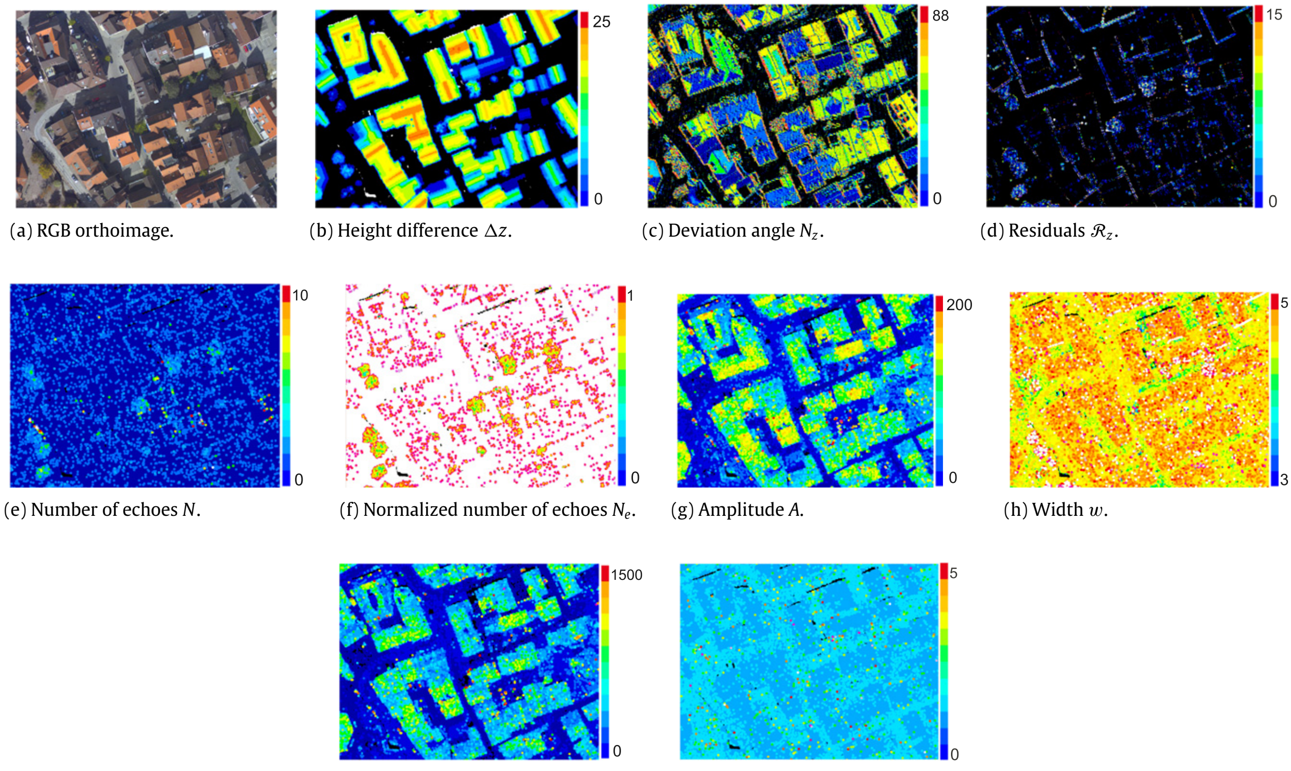
The feature vector f_v is defined as follows with three optical features R, G and B, five multi-echo lidar features and four full-waveform lidar features.

$$f_v = [R\ G\ B;\ \Delta z\ N_z\ \mathcal{R}_z\ N\ N_e;\ A\ w\ \sigma\ \alpha]^T. \qquad (2)$$

Table 1 summarizes the expected lidar feature values for the different classes we consider on urban scenes.

Fig. 4 depicts the median feature values for each class. The values are processed from the training dataset. A logarithmic scale is used for more clarity on small values. Feature values that equal 0 do not appear on such graphics. This is the case, for instance for Δz, for both types of grounds, and R_z for all classes except vegetation.

At first glance, one can notice the most discriminative features for all classes such as R, B, Δz, N_z, A, σ. Features N_z, R_z, N_e better discriminate the vegetation class. Ground classes give generally the same values except for optical image bands, and for FW lidar features A and σ. The second analysis consists in observing the lidar features (Fig. 3). Visually, we can assess whether a feature is important for the classification or not. For instance, FW lidar features (A and σ, Fig. 3(g) and (i)) lead to homogeneous values for buildings, hence minimizing this class intra-variance. FW lidar features are consequently well suited to the building class. Conversely, the echo width (w) values (Fig. 3(h)) are very noisy and do not help in discriminating the four classes visually. Moreover, when comparing the number of echoes and the normalized



(a) RGB orthoimage.

(b) Height difference Δz.

(c) Deviation angle N_z.

(d) Residuals $\mathcal{R}_z$.

(e) Number of echoes N.

(f) Normalized number of echoes N_e.

(g) Amplitude A.

(h) Width w.

(i) Cross-section σ.

(j) Echo shape α.

Fig. 3. 2D optical, multi-echo and full-waveform lidar features.

Table 1
Empirical values of lidar features for the four classes and class suitability.

Lidar	Feature	Building	Vegetation	Artificial ground	Natural ground	Suitability
ME	Δz	Variable	Variable	$\to 0$	$\to 0$	Ground vs off-ground
	N_z	[0°, 45°]	Variable	[0°, 10°]	[0°, 10°]	Ground
	$\mathcal{R}_z$	$\to 0$	High	$\to 0$	$\to 0$	Vegetation
	N	~ 1	>1	~ 1	1	Vegetation
	N_e	~ 1	≤ 1	~ 1	1	Vegetation
FW	A	Variable	Medium	Low	Variable	Artificial ground
	w	Medium	High	Variable	Variable	Vegetation
	σ	High	Medium	Medium	Variable	Building
	α	Variable	Variable	$\simeq\sqrt{2}$	$\simeq\sqrt{2}$	Natural ground

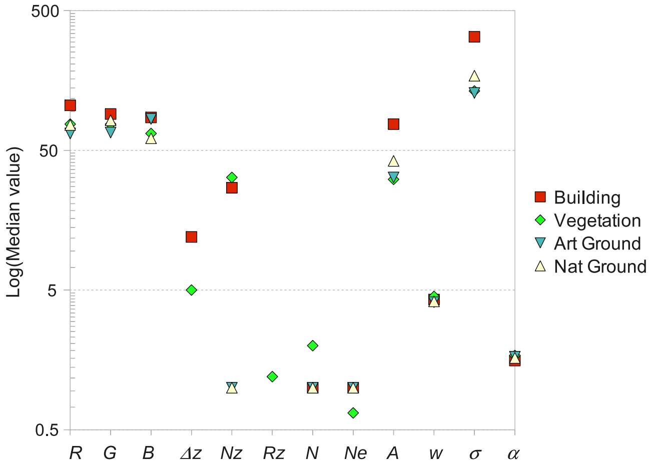


Fig. 4. Lidar features' median values per class.

number of echoes, the latter better highlights the vegetation. These observations will be confirmed by the feature importance measures in Section 5.1.

3. Random Forests classifier

Random Forests (RF) are a variant of bagging proposed by Breiman (Breiman, 2001). It is a decision-tree-based ensemble classifier that can achieve a classification accuracy comparable to boosting (Breiman, 2001), or SVM (Pal, 2005; Zhu, 2008). It does not overfit, runs fast and efficiently on large datasets such as lidar data. It does not require assumptions on the distribution of the data, which is interesting when different types or scales of input features are used. These outstanding properties make it suitable for remote sensing classification. It was successfully applied to multispectral data (Pal, 2005), multitemporal SAR images (Waske and Braun, 2009), hyperspectral data (Ham et al., 2005), or multi-source data (Gislason et al., 2006), where Landsat MSS and topographic data were used. Waske and Benediktsson (2007) applied it on SAR and multispectral images. For airborne multi-source classification using lidar and optical multispectral images, we showed in previous work that it achieves a classification accuracy comparable to SVM precision with a shorter training time (Chehata et al., 2009b).

In addition, the importance of each feature can be estimated during the training step. In this work, we exploit this property on

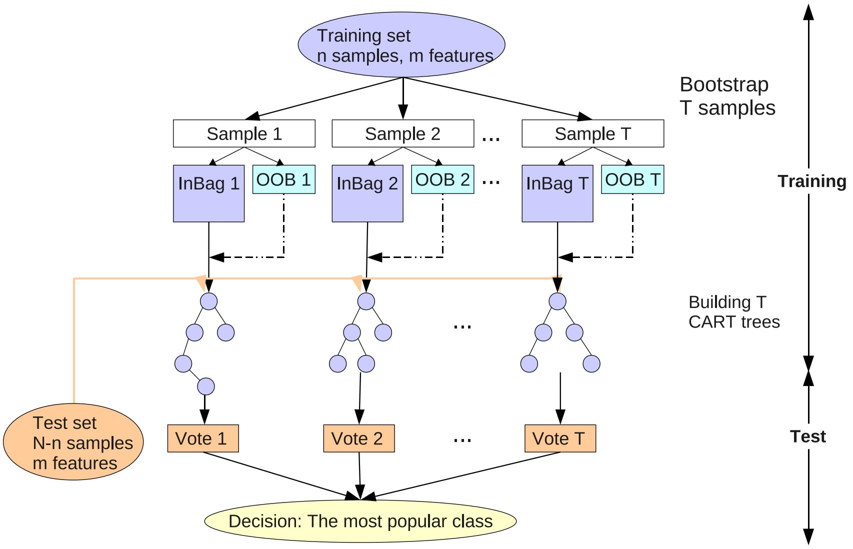


Fig. 5. The flow chart of Random Forests.

multi-source data in order to measure the relevance of airborne lidar and optical image features for classifying urban scenes.

3.1. Principle

Random Forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Breiman, 2001). In training, the algorithm creates T multiple bootstrapped samples of the original training data, and then builds a number of no pruned Classification and Regression Trees (CART) from each bootstrapped sample set. Only a randomly selected subset of the input features is considered to split each node of CART. The feature that minimizes the Gini impurity is used for the split (Breiman, 2001). For classification, each tree gives a unit vote for the most popular class at each input instance. The final label is determined by a majority vote of all trees. The RF classifier has two parameters: the number of trees T and the number of variables M randomly chosen at each split. Breiman's Random Forests error rate depends on two parameters: the correlation between any pair of trees and the strength of each individual tree in the forest. Increasing the correlation increases the forest error rate while increasing the strength of the individual trees decreases this misclassification rate. Reducing M reduces both the correlation and the strength. M is often set to the square root of the number of inputs (Breiman, 2001).

A global scheme highlighting the main steps of Random Forests is depicted in Fig. 5.

When the training set for a particular tree is drawn by sampling with replacement, about one-third of the cases are left out of the sample set. These samples are called Out-of-Bag (OOB) data (*cf.* Fig. 5) and are used to estimate the feature importance as detailed hereby. In the following, we denote by $\mathcal{B}^{(t)}$ the In-Bag samples for a tree t and by $\mathcal{B}^{c(t)}$ the complementary samples, *i.e.*, the OOB data for the tree t.

3.2. Feature importance measure

Aside from classification, Random Forests provide a measure of feature importance that is processed on OOB data and is based on the permutation importance measure (Breiman, 2001). The importance of a feature f can be estimated by randomly permuting all the values of this feature in the OOB samples for each tree. This follows the idea that a random permutation of a feature mimics the absence of that feature from the model. The measure of feature importance is the difference between prediction accuracy (*i.e.*, the number of observations correctly classified) before and after permuting feature f, averaged over all the trees (Breiman, 2001). A high prediction accuracy decrease denotes the importance of that feature. Suppose that the training samples consist of pairs of the form (x_i, l_j) where x_i is an instance, and l_j its true label. The importance of a feature f per tree t is computed as follows:

$$FI^{(t)}(f) = \frac{\sum\limits_{x_i\in\mathcal{B}^{c(t)}} I(l_j = c_i^{(t)})}{|\mathcal{B}^{c(t)}|} - \frac{\sum\limits_{x_i\in\mathcal{B}^{c(t)}} I(l_j = c_{i,\pi_f}^{(t)})}{|\mathcal{B}^{c(t)}|} \qquad (3)$$

where $\mathcal{B}^{c(t)}$ corresponds to OOB samples for a tree t, with $t \in \{1, \ldots, T\}$. $c_i^{(t)}$ and $c_{i,\pi_f}^{(t)}$ are the predicted classes for sample x_i before and after permuting the feature f respectively. Note that $FI^{(t)}(f) = 0$, if feature f is not in tree t. The importance score for a feature f is then computed as the mean importance over all trees:

$$FI(f) = \frac{\sum\limits_{T} FI^{(t)}(f)}{T} \qquad (4)$$

where T is the number of trees.

3.3. Margin definition

The margin concept of ensemble learning methods was first proposed by Schapire et al. (1998) to explain the success of boosting type algorithms. The concept was then generalized to analyze other types of ensemble classifiers (Breiman, 2001). Suppose that the training samples consist of pairs of the form (x_i, l_j), where x_i is an instance and l_j its true label. The margin m_i of instance x_i is computed as follows (Tang et al., 2006):

$$m_i = \mathrm{margin}(x_i, l_j) = \frac{v_{i,l_j} - \sum\limits_{c\neq l_j} v_{i,c}}{\sum\limits_{c} v_{i,c}} \qquad (5)$$

where v_{i,l_j} is the number of votes for the true class l_j, and $v_{i,c}$ is the number of votes for any class c with $c \neq l_j$. Hence, the margin is given by the difference between the fraction of classifiers voting correctly and incorrectly. It measures the strength of the vote. The margin ranges from -1 to $+1$. The positive margin value

Table 2
2D training and test samples. Class proportions show a highly imbalanced dataset.

Class	Training samples	Test samples	Proportion (%)
Building	17 617	71 398	46
Vegetation	1 616	6 752	4
Artificial ground	18 671	75 283	48
Natural ground	860	3 463	2
Total samples	38 764	156 896	100

of a sample indicates that this sample has been correctly classified, whereas a negative value means that the sample has been misclassified. The larger the margin, the more the confidence in the classification. A value close to 0 indicates a low confidence in the classification. Several studies have shown that the generalization performance of an ensemble classifier is related to the distribution of its margins on the training samples. Schapire et al. (1998) proved that achieving a larger margin on the training set leads to an improved bound on the generalization error of the ensemble.

In the following, the margin is exploited as a measure of confidence of the RF classifier. We use it to assess the relevance of input features.

4. Test dataset

The data is composed of georeferenced airborne lidar and multispectral *RGB* image. The lidar data acquisition was carried out with the RIEGL LMS-Q560 system over the city of Biberach (Germany). The main technical characteristics of this sensor are presented by Mallet and Bretar (2009). The lidar point cloud has a point density of approximatively 2.5 pts/m^2 with a footprint size of 0.25 m. The orthophotography has been captured with an Applanix DSS 22M device. Its resolution is 0.25 m and dimensions are 640 $*$ 485 pixels. To compute 2D lidar features, the following parameters were used: $c = 0.25$ m and $r_{\min} = 0.75$ m. This r value is chosen to provide the minimal number of lidar points (5 pts) in a 3D neighborhood.

The number of available reference samples is 195 660. 20% of randomly selected samples are used as a training set (*cf.* Table 2). One can observe that dense urban scenes are characterized by highly imbalanced classes: building and artificial ground are major classes while vegetation and natural ground are minor classes. The ground truth is processed manually, based on an oversegmentation of the orthoimage.

5. Results and discussion

The Random Forests implementation software by Leo Breiman and A. Cutler (http://www.r-project.org) was used in the experiments. Underlying parameters have been fixed to $M = \sqrt{n}$, where n is the number of features, and the number of trees T was set to 100.

Feature importance and margin confidence results will be presented in the first part, and then discussed with regard to physical lidar and multispectral image properties. The contribution of the considered features is detailed and explained not only for all classes, but also per class.

5.1. Feature importance results

To compute the feature importance, a balanced training set (3000 samples per class) is used to avoid biases due to a small number of samples of vegetation and natural ground classes in urban scenes. It is essential to select the best features for minor classes. A variable importance estimate for the training data is depicted in Fig. 6. The first three features are the optical components R, G, and B, whereas the latter are lidar features. The decrease permutation accuracy is averaged over all trees for all classes.

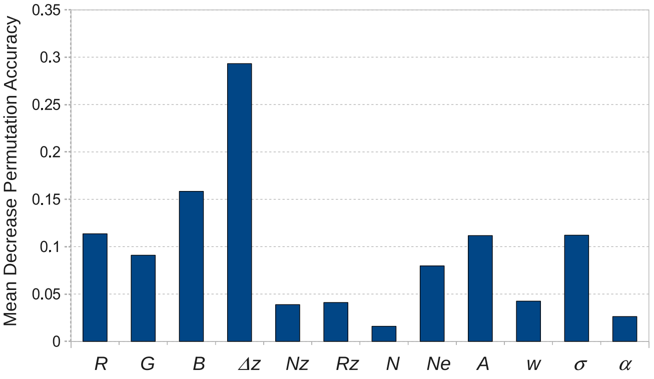


Fig. 6. Variable importance by mean decrease permutation accuracy. Balanced training data with 3000 samples per class.

From Fig. 6, it appears that the most relevant features for all classes are the height difference Δz, red and blue channels R, B, the echo amplitude A, and the echo cross-section σ. This leads to the following optimal feature vector: $[\Delta z, R, B, A, \sigma]$ that includes both optical image and FW lidar features as shown by Chehata et al. (2009b). This preliminary result will be confirmed in the following sections based on the margin theory.

5.2. Margin results

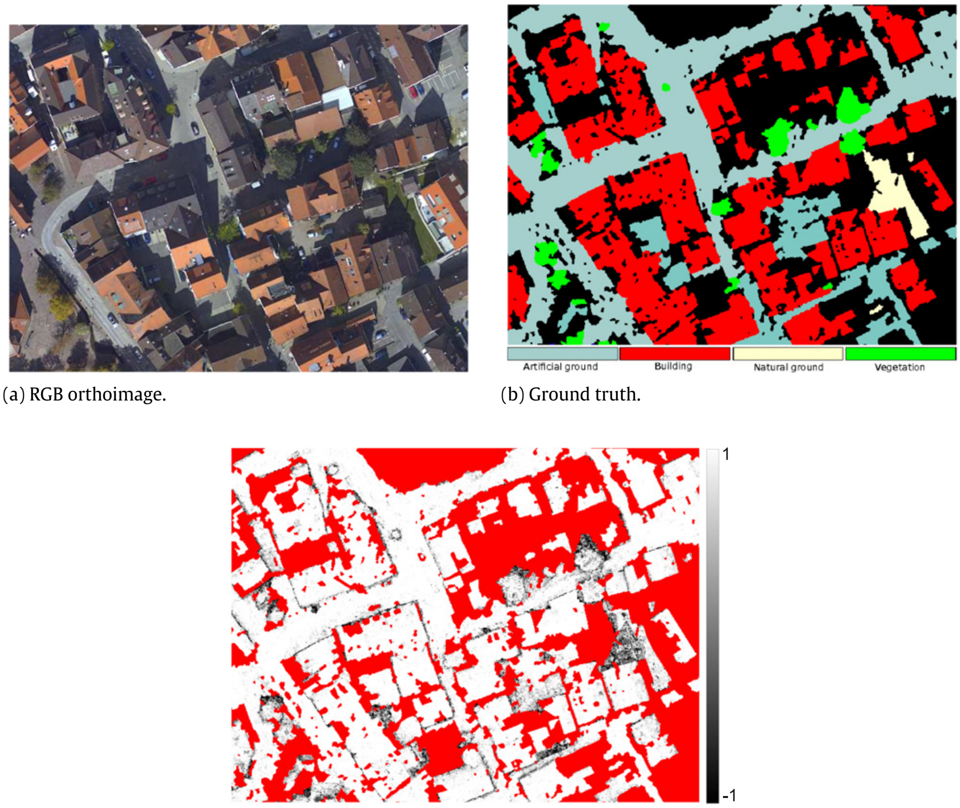
The margin is a confidence measure of predictions which helps in evaluating the classification quality and the RF classifier strength. Fig. 7 depicts the study area, the ground truth and the training margin image. This image was produced in the training step. All data are used as training samples just to illustrate the training margins.

The margin values are significantly higher at the center of classes in the feature space whereas the smaller margins correspond mainly to the class boundaries or to noise. In our case, class boundaries are located on building facades which are a transition between building and artificial ground classes, and also on vegetation boundary pixels that mix artificial ground and vegetation information. In fact, laser pulses can penetrate vegetation and reach the ground underneath. Consequently, for these areas, pixels combine various classes which make them harder to classify, leading to a low margin value. Moreover, shadowed areas are likely to have low margins when using R and B channels which have uniform irrelevant values in these areas. This is the case of natural ground points on the right of Fig. 7, or in urban corridors. Finally, higher margin values are located, in 2D space, at the center of buildings and artificial grounds as these classes are well discriminated using the five selected features.

5.3. Margin and classification confidence

Fig. 8 illustrates the test data margin histograms for well classified and misclassified samples.

For well classified samples, 88% of samples have a good classification confidence (margin ≥ 0.7). For misclassified samples, the margin distribution is more scattered and the confidence varies. Indeed, there are several sources of errors such as noise samples, class boundaries or even ground-truth errors. Considering the extreme case for misclassified samples with a high confidence (margin $= -0.9$), this is probably due to errors in the ground truth.



(a) RGB orthoimage.

(b) Ground truth.

(c) Margin image; five best features.

Fig. 7. Margin image based on the five best features — Random Forests classification ($T = 100$, $M = 2$). Black regions on the ground-truth image are not labelled and they correspond to the red regions in the margin image.

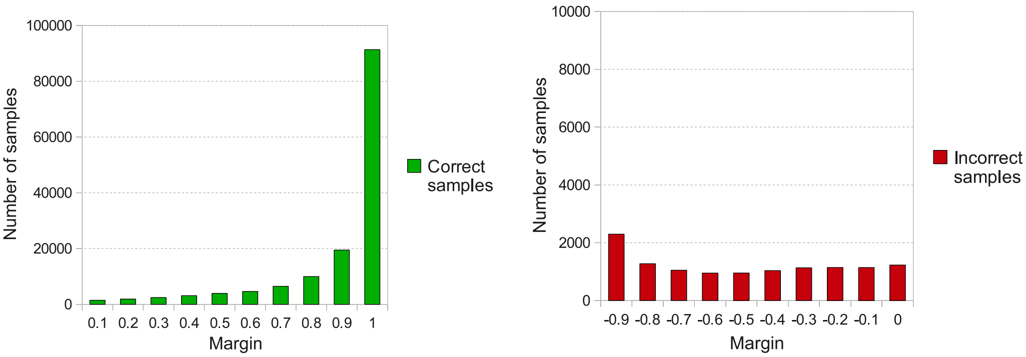


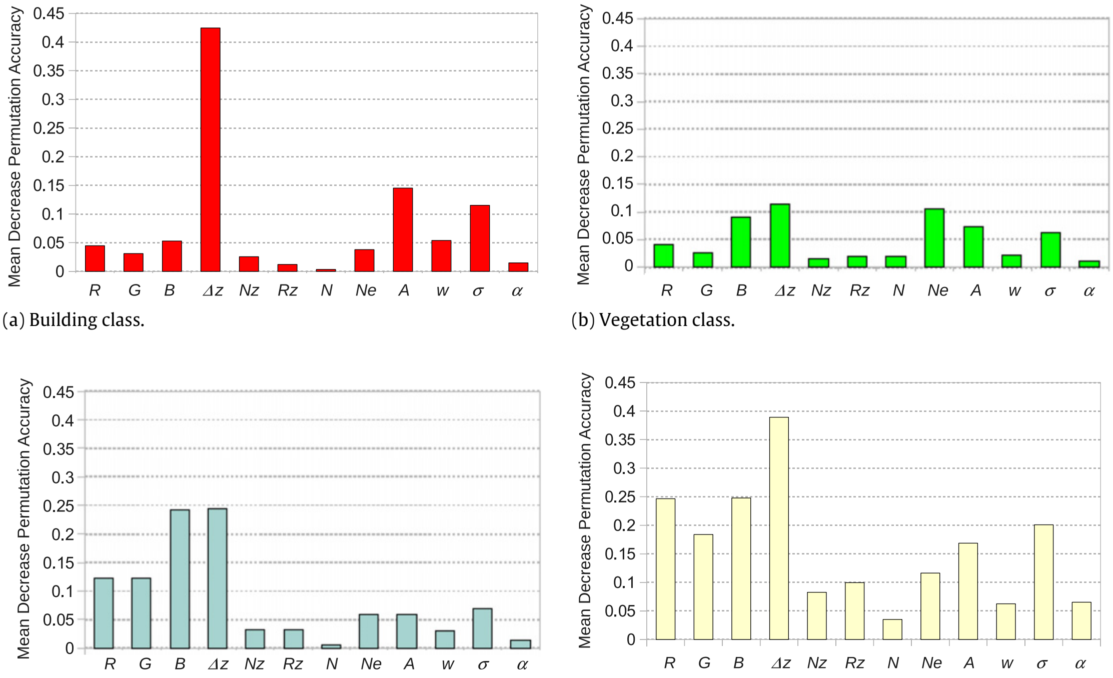
Fig. 8. Margin histograms for well classified and misclassified test data.

Table 3
Confusion matrix for test data using a RF classifier with the five best features, 100 trees and 2 split variables. Global error rate = 5.03%.

Obs.	Pred.				
	Building	Vegetation	Artificial ground	Natural ground	Omission error (%)
Building	69 128	242	1986	42	3.2
Vegetation	468	4882	1344	58	27.7
Artificial ground	1817	589	72 559	318	3.6
Natural ground	59	10	956	2438	29.6
Commission error (%)	3.3	14.7	5.7	14.6	5.0

5.4. Classification results

The Random Forests classifier is run with the five best selected features. The confusion matrix is given in Table 3. The training dataset is highly imbalanced with two major classes (building and artificial ground) that are more than ten times larger than vegetation and natural ground classes. We can notice that the artificial ground and buildings are well classified. However, the



(a) Building class.

(b) Vegetation class.

(c) Artificial ground class.

(d) Natural ground class.

Fig. 9. Feature importance per class using mean decrease permutation accuracy.

algorithm has more difficulties in classifying natural ground and vegetation which suffer from smaller training sets. Higher sources of errors are highlighted in Table 3.

Errors essentially occur between building and artificial ground since (1) these two classes have similar colors on optical images and the presence of shadows increases these errors, and (2) lidar features are ambiguous for building facades that are transitions between both classes. In addition, vegetation classes can be confused with artificial ground classes as the laser pulse can reach the ground under sparse vegetation. This confusion is increased by 2D lidar data interpolation. Finally, classification errors may occur between natural and artificial grounds since ME lidar features do not allow discriminating them.

Errors correspond to small margin values. They are depicted in Fig. 7.

5.5. *Feature relevance for urban scenes*

The feature relevance for urban scenes can be studied by the mean decrease permutation accuracy for all classes (*cf.* Fig. 6). From this figure, the height difference Δz appears to be the most important feature for all classes. It corresponds to the only topographic feature which helps in discriminating between ground and off-ground classes. Moreover, the study area is essentially composed of red brick rooftops and artificial ground which have high reflectances in the red and blue channels, respectively, leading to a high importance of both optical features. On the contrary, the number of echoes N is the less important feature. Indeed, this feature is not discriminating since there is one echo for all building rooftops and grounds, and it varies for vegetation. The other features will be discussed per class in the following section.

5.6. *Feature relevance per class*

Fig. 9 depicts the feature importance for each class. For the building class (Fig. 9(a)), the most important variables are Δz and two FW lidar features: A and σ. The latter features induce high values for buildings as explained in Section 2. Conversely, *RGB* channels are less important as they are not homogeneous on rooftops due to shadows as shown in Fig. 3(a). Both features that are related to the planarity of the local neighborhood (N_z, $\mathcal{R}_z$) show a low importance since they are very sensitive to the slope. These features are not homogeneous for the building class. However, they should be relevant for roof segmentation. As for the echo features, building facades lead to multiple echoes which explains the relative importance of the normalized number of echoes, N_e.

Regarding the vegetation class (Fig. 9(b)), in addition to the height difference, the normalized number of echoes N_e is very discriminating. A high number of echoes is specific to vegetation. Moreover, the normalized number of echoes is more important than the number of echoes since the values of the normalized number of echoes are bounded between 0 and 1 which is more appropriate for classification. Only one spectral band B, and FW lidar features A and σ are important. The latter features exhibit low values for vegetation as presented in Section 2.

As for multispectral bands, the spectral band G does not seem to be important in this case since it does not allow distinguishing vegetation from natural ground.

However, due to the laser properties, this class is composed of both vegetation and artificial ground information. Consequently, feature importances are more dispersed and height difference appears to be less important than for other classes.

The artificial ground class (Fig. 9(c)) shows more importance in multispectral *RGB* channels compared to lidar data. Indeed, due to urban corridors and building facades, points belonging to this class may also belong to the building class. Due to this confusion, height difference is less important.

As for the natural ground class (Fig. 9(d)), the feature importances are more dispersed between topographic, radiometric, and some lidar features which makes the classification more complex (*cf.* Table 3).

The echo shape feature seems to be more important for this class as low values of α correspond to the natural ground class (*cf.* Table 1). However, this feature is not much relevant for the

Table 4
Classification accuracy per class with respect to type and number of features.

Features	Building	Vegetation	Artificial ground	Natural ground	Total accuracy
RGB (3)	0.80	0.39	0.89	0.43	0.82
ME lidar (5)	0.96	0.50	0.95	0.01	0.92
FW lidar (4)	0.95	0.69	0.95	0.55	0.93
ME + FW lidar (9)	0.97	0.74	0.96	0.37	0.94
Selected features (5)	0.97	0.72	0.96	0.70	0.95
All features (12)	0.97	0.75	0.97	0.72	0.96

Table 5
Mean margin per class with respect to type and number of features.

Features	Building	Vegetation	Artificial ground	Natural ground	Mean margin
RGB (3)	0.55	−0.35	0.70	−0.22	0.57
ME lidar (5)	0.86	−0.10	0.87	−0.94	0.78
FW lidar (4)	0.81	0.20	0.78	−0.01	0.75
ME + FW lidar (9)	0.89	0.28	0.82	−0.22	0.80
Selected features (5)	0.89	0.33	0.87	0.31	0.84
All features (12)	0.89	0.35	0.85	0.30	0.84

classification of dense urban scenes since natural ground is a minor class.

5.7. Multi-source feature analysis

In order to confirm the relevance of most important features, classification was run using multispectral image, ME lidar and FW lidar features separately, then using all lidar features, and finally combining optical image and lidar features. The accuracies and the margins are compared to those obtained using the five selected features.

5.7.1. Classification accuracy analysis

Table 4 sums up the accuracy results for all classes as well as per class. The results are averaged over ten classification runs. They confirm the importance of the joint use of multispectral images and lidar data for urban scene classification, and the contribution of FW lidar features which are both highlighted in the table.

When observing the extensive accuracy comparison (*cf.* Table 4), many points are highlighted.

1. The *RGB* bands result in the lowest global classification accuracy and fail to correctly classify the vegetation class. This is due to image sensitivity to the illumination angles (building rooftops), and due to the presence of shadows in dense urban scenes.
2. ME lidar features are suited to building and artificial ground classification. However, they do not allow discriminating between both types of grounds since ME lidar data only provide spatial information. Consequently, natural ground points are misclassified as artificial ground class which is the major class. As for vegetation, only an accuracy of 50% is achieved due to the laser properties in these areas (*cf.* Section 5.4).
3. FW lidar features give better results than ME lidar without using Δz. In fact, the four FW lidar features are complementary, as shown in Table 1. The amplitude A discriminates artificial ground by low values. The echo width w is well suited to vegetation. The echo cross-section σ discriminates between building and artificial ground leading to high and low values respectively, and finally α is important for the natural ground class (*cf.* Fig. 9(d)). This confirms the relevance of FW lidar data for classifying urban scenes.
4. Using all nine lidar features enhances the results especially for buildings and vegetation. Both classes are considered as off-ground which is better discriminated using Δz. However, the results are worse for the natural ground class. As shown in Fig. 9(d), the variable importance is dispersed for this class. Using all lidar features disturbs the training model which results in a lower accuracy.
5. The most relevant five features maintain a satisfactory classification accuracy of 95% while reducing the number of features. When compared to all features' results, accuracies mainly differ for minor classes. As shown in Fig. 9(b) and (d), the feature importance is dispersed. Reducing the number of features affects essentially these minor classes.

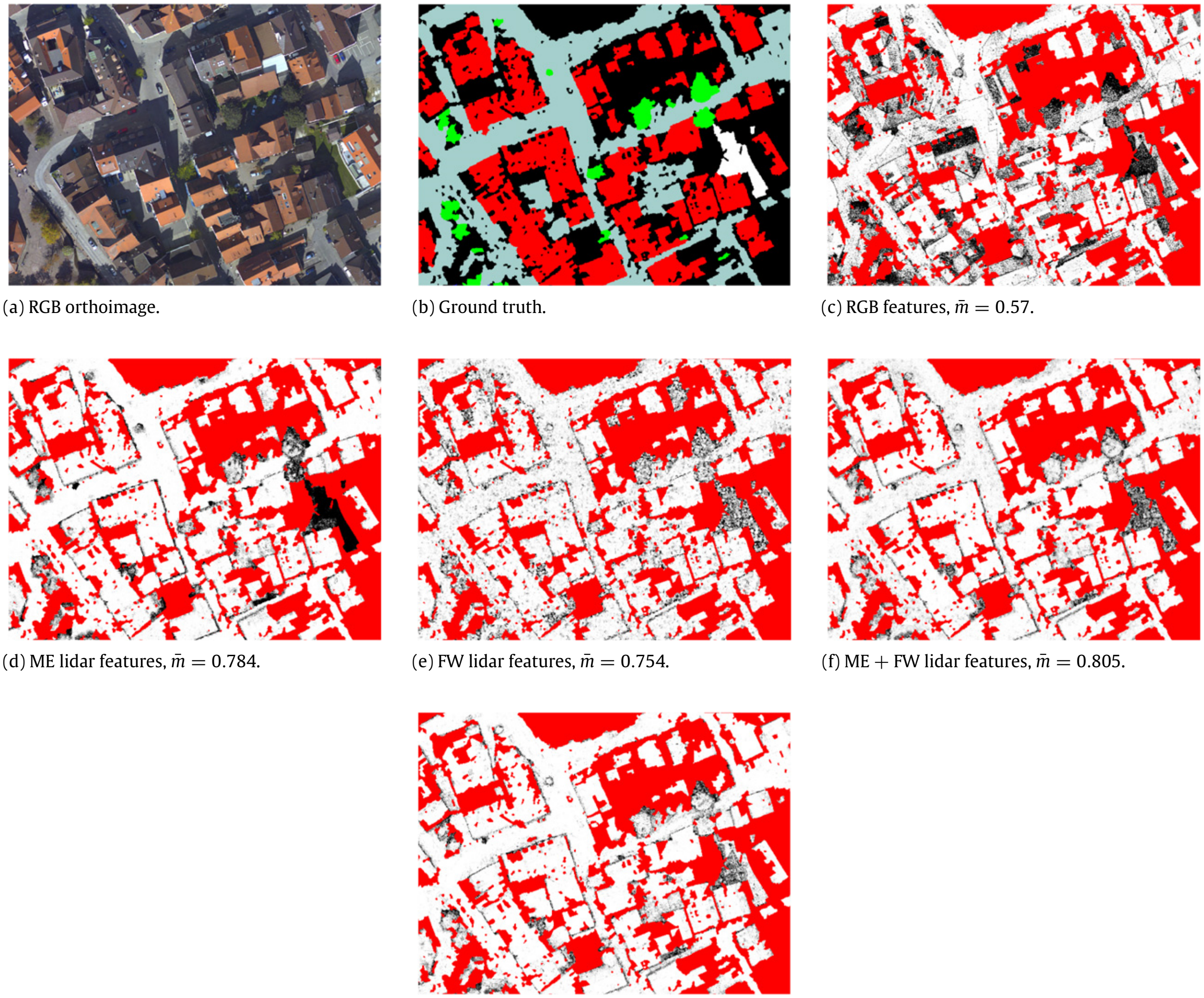
5.7.2. Margin analysis

Table 5 compares the mean margin value for different input features. One can observe that the highest margin is obtained with the five selected features. The higher the margin, the stronger the ensemble classifier. A high negative margin value indicates that most trees voted for wrong classes. This may be due to the non-suitability of the input features for the true class or to noisy samples. The joint interpretation of Tables 4 and 5 shows that the RF classifier with the best five features has the highest positive mean margin while keeping a good classification accuracy. This confirms that the five selected features are relevant for urban scene classification. The confidence is even improved in comparison with the RF classifier involving all features for all classes except for vegetation. The former classifier maximizes the minimal margins for these classes. However, the latter classifier performs better for vegetation since two of the remaining features are well suited to vegetation (N_e and R_z, *cf.* Fig. 4).

We can notice that using only ME lidar features for the natural ground class leads to a high negative margin and a poor classification accuracy. The margin value leads to a high confidence in misclassification; in this case, the input features are not suited to the class of interest. Indeed, ME lidar features are spatial features which do not discriminate between both types of grounds.

The results show that FW lidar features are required to distinguish between artificial and natural grounds. In addition, they significantly improve the classification of vegetation. The global accuracy is enhanced in comparison with ME lidar features. The minimal margins for vegetation and natural ground are improved since FW lidar features better characterize the physical properties of objects, but the confidence still remains low with values around 0. However, with FW lidar features, the confidence is lower for major classes due to the non-use of Δz which better discriminates between buildings and artificial ground.

When combining ME and FW lidar features, the confidence is enhanced for all classes except natural ground. For the latter class, the ME lidar features are not suitable as explained before. So they introduce a kind of noise for this class. The joint use of both lidar acquisition modes highly improves the classification accuracy of the vegetation class.



(a) RGB orthoimage.

(b) Ground truth.

(c) RGB features, $\bar{m} = 0.57$.

(d) ME lidar features, $\bar{m} = 0.784$.

(e) FW lidar features, $\bar{m} = 0.754$.

(f) ME + FW lidar features, $\bar{m} = 0.805$.

(g) Five selected features, $\bar{m} = 0.844$.

Fig. 10. Margin image comparison depending on multi-source features.

Fig. 10 depicts the corresponding margin images with *RGB*, ME lidar, FW lidar, joint use of ME and FW lidar, and the best five features. With multispectral *RGB* features (Fig. 10(c)), one can observe that margin values are low on shadowed rooftops and grounds leading to the lowest mean margin $\bar{m}$. When using ME lidar features (Fig. 10(d)), the margin values are higher especially for major classes (building and artificial ground). We can see the natural ground class in black color due to the non-suitability of features to this class. For the FW lidar features' margin (Fig. 10(e)), the image presents globally a "salt and pepper" effect which denotes a lower confidence for major classes and still lower values for minor classes even if they are enhanced (*cf.* Table 5). This is due to noisy initial FW features. However, the global accuracy is improved (*cf.* Table 4). The joint use of ME and FW lidar features enhances the classification accuracy for all classes except for natural ground due to ME lidar properties. The confidence measure is improved for building and vegetation classes (*cf.* Table 5). This can be observed from the corresponding image margin (Fig. 10(f)). Indeed, among FW lidar features, the echo amplitude A is well suited to building, whereas echo cross-section σ is adapted for vegetation.

Finally, the mean margin of the RF classifier with five features (Fig. 10(g)) appears less noisy. Margin values are higher for all classes. Small margin values are mostly located on building facades and shadowed areas of vegetation and natural ground.

6. Conclusion

In this paper, we have studied the relevance of multi-source optical image and lidar data features for urban scene classification. Lidar features are composed of multi-echo and full-waveform features. The Random Forests classifier is used to assess both the feature importance measures, and the classifier strength. The permutation accuracy criteria have been used to study the importance of each feature for the whole urban scene and for each class. Through our experiments, several observations have been made: (1) the most significant feature is topographic, the relative height of a lidar point; (2) the best five feature vector is $[\Delta z, B, R, A, \sigma]$ that is composed of two optical image channels, one topographic ME lidar feature, and two FW lidar features. This result shows the relevance of a joint use of airborne lidar data and optical image features; (3) the contribution of FW lidar features is demonstrated in comparison with ME lidar features. The results are enhanced especially when the spatial information is insufficient (*i.e.* natural ground); (4) some features appear to be specific to a particular class such as the normalized number of echoes N_e for the vegetation.

Moreover, the margin theory has been introduced as a measure of confidence and to assess the strength of a RF classifier. We observed that: (1) confidence for the RF classifier with *RGB* features is very low especially in shadowed areas; (2) ME lidar provides a high confidence for building and artificial ground and a very poor confidence for natural ground since the related features are only topographic or geometric; (3) FW lidar enhances the confidence for building and vegetation classes. The RF classifier with the five selected features has the highest mean margin while achieving a good accuracy. This confirms that these features are the most discriminating for urban scene classification. A global classification accuracy of 95% is achieved. However, in dense urban scenes, the data are highly imbalanced, and, therefore the minor classes (vegetation and natural ground) are harder to classify leading to accuracies of respectively 72% and 70%. This result could be improved by a two-pass Random Forests classifier more adapted to minor classes and involving low margins. In addition, using an adaptive 2D neighborhood for lidar features, or radiometric equalization of multispectral image should give higher importance to the corresponding features. This study also showed that some features are not of significance to extract the four urban classes such as planarity related features (N_z, $\mathcal{R}_z$). They should be more useful for roof building segmentation.

The proposed feature importance framework can be applied to select the appropriate features for classification, segmentation or multi-source fusion applications. It can also be applied to remote sensing data such as hyperspectral data which involve hundreds of features, as well as multi-source, multi-sensor data. Finally, this work can also benefit other application fields such as biomedical or computer vision.

References

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

Brenner, C., 2010. In: Vosselman, G., Maas, H.-G. (Eds.), Airborne and Terrestrial Laser Scanning. Whittles Publishing, Dunbeath, Scotland, UK.

Carlberg, M., Gao, P., Chen, G., Zakhor, A., 2009. Classifying urban landscape in aerial LiDAR using 3D shape analysis. In: Proceedings of the IEEE International Conference on Image Processing. IEEE, Cairo, Egypt, pp. 1701–1704.

Charaniya, A., Manduchi, R., Lodha, S., 2004. Supervised parametric classification of aerial LiDAR data. In: Proceedings Real-Time 3D Sensors and their Use Workshop, in Conjunction with IEEE CVPR. Washington, DC, USA. 27 June–2 July (on CD-ROM).

Chauve, A., Mallet, C., Bretar, F., Durrieu, S., Pierrot-Deseilligny, M., Puech, W., 2007. Processing full-waveform LiDAR data: modelling raw signals. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part 3/W52), 102–107.

Chehata, N., Guo, L., Mallet, C., 2009a. Airborne LiDAR feature selection for urban classification using random forests. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 39 (Part 3/W8), 207–212.

Chehata, N., Guo, L., Mallet, C., 2009b. Contribution of airborne full-waveform LiDAR and image data for urban scene classification. In: Proceedings IEEE International Conference on Image Processing (ICIP). IEEE, Cairo, Egypt, pp. 1669–1672.

Gislason, P., Benediktsson, J., Sveinsson, J., 2006. Random forests for land cover classification. Pattern Recognition Letters 27 (4), 294–300.

Gross, H., Jutzi, B., Thoennessen, U., 2007. Segmentation of tree regions using data of a full-waveform laser. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (Part (3/W49A)), 57–62.

Haala, N., Brenner, C., 1999. Extraction of buildings and trees in urban environments. ISPRS Journal of Photogrammetry and Remote Sensing 54 (2–3), 130–137.

Ham, J., Chen, Y., Crawford, M., Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 43 (3), 492–501.

Höfle, B., Hollaus, M., 2010. Urban vegetation detection using high density full-waveform airborne LiDAR data—combination of object-based image and point cloud analysis. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (Part 7B), 281–286.

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

Mallet, C., Bretar, F., Soergel, U., 2008. Analysis of full-waveform LiDAR data for classification of urban areas. Photogrammetrie Fernerkundung GeoInformation (PFG) 2008 (5), 337–349.

Matei, B., Sawhney, H., Samarasekera, S., Kim, J., Kumar, R., 2008. Building segmentation for densely built urban regions using aerial LiDAR data. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Anchorage, AK, USA, pp. 1–8.

Matikainen, L., Hyyppä, J., Hyyppä, H., 2003. Automatic detection of buildings from laser scanner data for map updating. International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences 33 (Part 3/W13), 218–224.

Melzer, T., 2007. Non-parametric segmentation of ALS point clouds using mean shift. Journal of Applied Geodesy 1 (3), 159–170.

Pal, M., 2005. Random forest classifier for remote sensing classification. International Journal of Remote Sensing 26 (1), 217–222.

Poullis, C., Yu, S., 2009. Automatic reconstruction of cities from remote sensor data. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Miami, FL, USA, pp. 2775–2782.

Reitberger, J., Schnorr, C., Krzystek, P., Stilla, U., 2009. 3ṣegmentation of single trees exploiting full waveform LiDAR data. ISPRS Journal of Photogrammetry and Remote Sensing 64 (6), 561–574.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2007. Building detection by fusion of airborne laser scanner data and multi-spectral images: performance evaluation and sensitivity analysis. ISPRS Journal of Photogrammetry and Remote Sensing 62 (2), 135–149.

Rutzinger, M., Höfle, B., Hollaus, M., Pfeifer, N., 2008. Object-based point cloud analysis of full-waveform airborne laser scanning data for urban vegetation classification. Sensors 8 (8), 4505–4528.

Schapire, R., Freund, Y., Bartlett, P., Lee, W., 1998. Boosting the margin: a new explanation for the effectiveness of voting methods. The Annals of Statistics 26 (5), 1651–1686.

Secord, J., Zakhor, A., 2007. Tree Detection in urban regions using aerial LiDAR and image data. IEEE Geoscience and Remote Sensing Letters 4 (2), 196–200.

Sithole, G., Vosselman, G., 2004. Experimental comparison of filter algorithms for bare-earth extraction from airborne laser scanning point clouds. ISPRS Journal of Photogrammetry and Remote Sensing 59 (1–2), 85–101.

Tang, E.K., Suganthan, P.N., Yao, X., 2006. An analysis of diversity measures. Machine Learning 65 (1), 247–271.

Wagner, W., Hollaus, M., Briese, C., Ducic, V., 2008. 3D vegetation mapping using small-footprint full-waveform airborne laser scanners. International Journal of Remote Sensing 29 (5), 1433–1452.

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

Waske, B., Benediktsson, J., 2007. Fusion of support vector machines for classification of multisensor data. IEEE Transactions on Geoscience and Remote Sensing 45 (12), 3858–3866.

Waske, B., Braun, M., 2009. Classifier ensembles for land cover mapping using multitemporal sar imagery. ISPRS Journal of Photogrammetry and Remote Sensing 64 (5), 450–457.

Xu, G., Zhang, Z., 1996. Epipolar Geometry in Stereo, Motion and Object Recognition. Kluwer Academic Publishers, Boston, MA, USA.

Zhou, Q.-Y., Neumann, U., 2009. A streaming framework for seamless building reconstruction from large-scale aerial LiDAR data. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Miami, FL, USA, pp. 2759–2766.

Zhu, M., 2008. Kernels and ensembles: perspectives on statistical learning. The American Statistician 62 (2), 97–109.