Semantic segmentation of forest stands of pure species combining airborne lidar data and very high resolution multispectral imagery

Clément Dechesne, Clément Mallet, Arnaud Le Bris, Valérie Gouet-Brunet

*Univ. Paris-Est, LASTIG MATIS, IGN, ENSG, F-94160 Saint-Mande, France*

**Abstract**

Forest stands are the basic units for forest inventory and mapping. Stands are defined as large forested areas (e.g., \( \geq 2 \text{ha} \)) of homogeneous tree species composition and age. Their accurate delineation is usually performed by human operators through visual analysis of very high resolution (VHR) infra-red images. This task is tedious, highly time consuming, and should be automated for scalability and efficient updating purposes. In this paper, a method based on the fusion of airborne lidar data and VHR multispectral images is proposed for the automatic delineation of forest stands containing one dominant species (purity superior to 75\%). This is the key preliminary task for forest land-cover database update. The multispectral images give information about the tree species whereas 3D lidar point clouds provide geometric information on the trees and allow their individual extraction. Multi-modal features are computed, both at pixel and object levels: the objects are individual trees extracted from lidar data. A supervised classification is then performed at the object level in order to coarsely discriminate the existing tree species in each area of interest. The classification results are further processed to obtain homogeneous areas with smooth borders by employing an energy minimum framework, where additional constraints are joined to form the energy function. The experimental results show that the proposed method provides very satisfactory results both in terms of stand labelling and delineation (overall accuracy ranges between 84\% and 99\%).

*Keywords:* Lidar, multispectral imagery, fusion, feature selection, supervised classification, energy minimisation, regularisation, forest stand delineation, tree species.

1. Introduction

1.1. Motivation

Fostering information extraction in forested areas from remote sensing data, in particular at the stand level, is driven by two main goals: statistical inventory and mapping. Forest stands are the basic units for subsequent analysis and can be defined in terms of tree species or tree maturity. From a remote sensing point of view, the delineation of the stands is a segmentation problem. In statistical national forest inventory (NFI), an automated and accurate tree segmentation is needed in order to extract tree level features (basal area, dominant tree height, etc., (Means et al., 2000; Kangas and Maltamo, 2006)). However, the tree level is not the only reliable level of analysis for forest studies, a larger scale (e.g., forest stands) is interesting in order to extract reliable and statistically meaningful features and to provide an input for multi-source statistical inventory. For land-cover mapping, this is highly helpful for forest database updating (Kim et al., 2009), whether the labels of interest are vegetated areas (e.g., deciduous/evergreen/mixed/non-forested), or, more precisely, the tree species. Most of the time in national forestry inventory institutes, for reliability purposes, each area is manually interpreted by human operators with very high resolution (VHR) geospatial images focusing on the infra-red channel (Kangas and Maltamo, 2006). This work is extremely time consuming and subjective (Wulder et al., 2008). Furthermore, in many countries, the wide variety of tree species (e.g., >20) significantly complicates the problem. The design of an automatic procedure based on remote sensing data would fasten such process. Additionally, the standard manual delineation procedure only takes into account the species, and few characteristics (alternatively height, age, stem density or crown closure), while an automatic method could offer more flexibility and would allow to combine characteristics extracted from all complementary data sources.

1.2. State of the art

The use of remote sensing data for the automatic analysis of forests has been growing in the last 15 years, especially with the synergistic use of airborne laser scanning (ALS) and optical VHR imagery (multispectral imagery and hyperspectral imagery) (Torabzadeh et al., 2014; White et al., 2016). In the large amount of literature in the field, only few papers focus on the issue of stand segmentation or delineation. They can be categorised with regard
to the type of data processed.

First, stand segmentation can be achieved with a single remote sensing source. A stand delineation technique using VHR airborne multispectral imagery is proposed in [Leckie et al. 2003]. The trees are extracted using a valley following approach and classified into 7 tree species (5 coniferous, 1 deciduous, and 1 non-specified) with a maximum likelihood classifier. A semi-automatic iterative clustering procedure is then introduced to generate the forest polygons.

A hierarchical and multi-scale approach for the identification of stands is adopted in [Hernando et al. 2012]. The data inputs were the 4 bands of an airborne 0.5 m orthoimage (Red, Green, Blue, and Near Infra-Red) allowing to derive the Normalised Difference Vegetation Index (NDVI). The stand mapping solution is based on the Object-Based Image Analysis concept. It is composed of two main phases in a cyclic process: first, segmentation, then classification. The first level consists in over-segmenting the area of interest and performing fine-grained land cover classification. The second level aims to transfer the vegetation type provided by a land cover geodatabase in the stand polygons, already retrieved from another segmentation procedure. The multi-scale analysis appears to have a significant benefit on the stand labelling but it is highly heuristic and requires a correct definition of the stand while we consider it is an interleaved problem.

A seminal stand mapping method using low density airborne lidar data is proposed in [Koch et al. 2009]. It is composed of several steps of feature extraction, creation and raster-based classification. Forest stands are created by grouping neighbouring cells within each class. Then, only the stands with a pre-defined minimum size are accepted. Neighbouring small areas of different forest types that do not reach the minimum size are merged together to an existing forest stand. The approach offers the advantage of detecting 15 forest types that match very well with the ground truth but to the detriment of simplicity: the flowchart has to be highly reconsidered to fit to other stand specifications. Additionally, the tree species discrimination is not addressed. The forest stand delineation proposed in [Sullivan et al. 2009] also uses low density airborne lidar still coupling an object-oriented image segmentation and a supervised classification procedure. Three features are computed and rasterized. The segmentation is performed using a region growing approach. Spatially adjacent pixels are grouped into homogeneous discrete image objects or regions. Then, a supervised discrimination of the segmented image
is performed using a Battacharya classifier, in order to determine the maturity of the stands. The tree species are ignored and the procedure requires a careful inspection of the raw data both for feature generation and model training.

Following the work of (Wulder et al., 2008) with IKONOS images, Quickbird-2 panchromatic images are used in (Mora et al., 2010) to automatically delineate forest stands. A standard image segmentation technique is used and the novelty mainly lies on the fact that its initial parameters are optimized with respect to NFI protocols. They show that meaningful stand heights can be derived, which are a critical input for various modelled inventory attributes.

The method proposed in (Eysn et al., 2012) aims to generate a forest mask (forested area label only) using low density airborne lidar. A Canopy Height Model (CHM) with a spatial resolution of 1 m is derived. The positions and heights of single trees are determined from the CHM using a local maximum filter, based on a moving window approach. Only detected positions with a CHM height superior to 3 m are considered. The crown radii are estimated using an empirical function. The three neighbouring trees are connected using a Delaunay triangulation applied to the previously-detected tree position. The crown cover is then calculated using the crown areas of three neighbouring trees and the area of their convex hull for each tree triple. The forest mask is derived from the canopy cover values. While this is not a genuine stand delineation method, this approach could be easily extended to a multi-class problem and enlightens the necessity of individual tree extraction even with limited point densities as a basis for the stand-level analysis.

A forest stand delineation also based on airborne lidar data is proposed in (Wu et al., 2014). Three features are first directly extracted from the point cloud. A coarse forest stand delineation is then performed on the feature image using the unsupervised Mean-Shift algorithm, in order to obtain under-segmented raw forest stands. A forest mask is then applied to the segmented image in order to retrieve forest and non-forest raw stands. It may create some small isolated areas, iteratively merged to their most similar neighbour until their size is larger than a user-defined threshold in order to produce big raw forest stands. They are then refined into finer level using a seeded region growing based on superpixels. The idea is to select several different superpixels in a raw forest stand and merge them. This method provides a coarse-to-fine segmentation with relatively large stands. The process was only applied on a small area of a forest in Finland, thus, general conclusions can not be drawn.
Secondly, several methods fusing various types of remote sensing data have also been developed. The analysis of the lidar and multispectral data is performed at three levels in (Tiede et al., 2004), following a given hierarchical nomenclature of classes in forested environments. The first level represents small objects (single tree scale, individual trees or small groups of trees) that can be differentiated by spectral and structural characteristics using a rule-based classification. The second level corresponds to the stand level. It is built using the same classification process which summarizes forest development phases by referencing to small scale sub-objects at level 1. The third level is generated by merging objects of the same classified forest-development into larger spatial units. The multi-scale analysis offers the advantage of alleviating the standard issue of individual tree crown detection and proposing development stage labels. Nevertheless, the pipeline is highly heuristic, under-exploits lidar data and significant confusion between classes are reported.

The automatic segmentation process of forests in (Diedershagen et al., 2004) is also supplied with lidar and VHR multispectral images. The idea is to divide the forests into higher and lower sections with lidar. An unsupervised classification process is applied to the two new images. The final stand delineation is achieved by segmenting the classification results with pre-defined thresholds. The segmentation results are improved using morphological operators such as opening and closing, which fill the gaps and holes at a specified extent. This method is efficient if the canopy structure is homogeneous and requires a strong knowledge on the area of interest. Since it is based on height information only, it cannot differentiate two stands of similar height but different species.

In (Leppänen et al., 2008) a stand segmentation technique for a forest composed of Scots Pine, Norway Spruce and Hardwood is defined. A hierarchical segmentation on the Crown Height Model followed by a restricted iterative region growing approach is performed on images composed of rasterized lidar data and Coloured Infra-Red images. The process was only applied on a limited area of Finland and prevents from drawing strong conclusions. However, the quantitative analysis carried out by the authors shows that lidar data can help to define statistically meaningful stands (here the criterion was the timber volume) and that multispectral images are inevitable inputs for tree species discrimination.
1.3. Contribution

With respect to the methods mentioned above, it appears that there are no forest stand segmentation method, based on tree species, that can satisfactorily handle a large number of classes (>5). It also appears that working at the object level (usually the tree level), in order to discriminate the tree species, produces better stand segmentation results in addition to better discriminate the species themselves (Hovi et al., 2016). Several methods for tree species classification at the tree level have been proposed in (Leckie et al., 2003), (Heinzel and Koch, 2012) and (Dalponte et al., 2015a) using lidar point clouds and/or multi/hyperspectral images. Although they are likely to be imprecise, resulting in a non-perfect tree species discrimination (up to 80% in most studies), they can be considered as a relevant hint for forest stand delineation: a coarse knowledge may be sufficient as initial guess. The same conclusion applies for individual tree crown extraction from lidar data.

In this paper, we propose a fully automatic method for species-based forest stand segmentation. The method is composed of three main steps; feature computation, vegetation type (mainly tree species) classification and regularisation. Features are first derived at the pixel and at the tree level. The most relevant ones are subsequently selected in a supervised way. The trees are extracted using a simple yet efficient method, since it appears to be sufficient for subsequent steps. A classification is performed at the tree level as it significantly improves the discrimination results (about 10% better than the pixel-based approach). This classification is then regularised through an energy minimisation process. The regularisation, which is performed with a standard graph-cut method, produces homogeneous vegetation type (mainly tree species) areas with smooth borders. Our contributions are two-fold:

- Such framework can be fed with specific constraints allowing to tailor the results to specific criteria (height, age, specie, maturity, density, ...).
- Here, the training set is automatically derived from an existing forest land-cover geodatabase. Specific attention is paid to the extraction of the most relevant training pixels, which is highly challenging with outdated and generalized vector databases.

This paper is an extension of (Dechesne et al., 2016) with additional methodological contribution (namely feature selection and automatic training set...
design), additional experiments and discussions. The method is now fully automatic, and has been designed for scalability purpose. Modifications have therefore been performed on the classification and the regularisation steps. The paper is structured as follows: the method is presented in Section 2. The employed dataset is presented in Section 3 while the experiments and results are discussed in Section 4. Finally, conclusions are drawn in Section 5 with a summary of the method and its forthcoming improvements.

2. Methods

The proposed method is composed of three main steps illustrated by the flowchart in Figure 1. First, 95 features are derived from the airborne lidar (ALS) point clouds and multispectral images (Section 2.1). They are computed at the pixel level, for the subsequent energy minimisation procedure. They are also computed at the object level (namely, trees, extracted from the lidar data); the features are more consistent at the tree level and improve the accuracy of the classifier, as stated in a previous study (Dechesne et al., 2016). Then, the classification of the vegetation type (mainly tree species) is performed using a standard Random Forest supervised classifier (Section 2.2). Finally, a smoothing regularisation process based on an energy minimisation framework is carried out in order to obtain the final segments (Section 2.3).

2.1. Feature computation

It is composed of 5 steps.

- A coarse individual tree extraction from the ALS point cloud is first performed;
- Lidar features are computed at the 3D point level;
- Multispectral features are generated at the pixel level;
- The lidar features are rasterized at the resolution of the multispectral image;
- Finally, an object-based feature map is created jointly benefiting from the pixel-based feature maps and the extracted trees.

A total of 95 features are derived. An automatic feature selection is then carried out to limit both model complexity and memory consumption. It highlights the complementarity and the relevance of the derived features.
Figure 1. Flowchart of the proposed method. Colour codes: ◆chestnut, ◆robinia, ◆deciduous oaks, ◆fir or spruce, ◆no data, ©IGN
2.1.1. **Tree extraction.**

Lidar data were first pre-processed: the point cloud was filtered in order to remove outliers and determine ground points. The ground points were extrapolated so as to derive a Digital Terrain Model (DTM) at 1 m resolution. Then, the DTM was subtracted from the point cloud since a normalized point cloud is necessary in order to perform our coarse tree delineation. Tree delineation has been extensively investigated in the literature, see for instance ([Dalponte et al., 2015b](#)) and ([Tochon et al., 2015](#)) for the latest developments. A large range of methods exist ([White et al., 2016](#)). It is well known there is no golden standard for individual tree crown delineation ([Kaartinen et al., 2012](#); [Wang et al., 2016](#)). Here, accurate delineation is not the aim of our proposal. Instead, a geometrically meaningful over-segmentation technique for object-based image analysis and feature computation is desired. Consequently, and for scalability purposes, a simple method is sufficient.

![Image 1](image1.png) ![Image 2](image2.png)

(a) Orthoimage of 1 km² ©IGN. (b) Sub-area with delineated trees.

Figure 2. Coarse tree delineation as a support for robust feature computation.

A coarse method is therefore adopted: the tree tops are first extracted from the lidar point clouds using a local maximum filter. A point is considered as a tree top when it has the highest height value within a 5 meter radius. Only the points above 3 meters are retained as it is a common threshold of the literature ([Eysn et al., 2012](#)), and appears to be highly discriminative in non-urban areas. Points belonging to a tree are obtained through two criteria. (i) If the height of a point within a 5 m radius is greater or equal than 80% the height of the closest tree top, it is aggregated to the tree top. (ii) If
the distance in the $(x,y)$ plane between an unlabelled point and the closest tree point is smaller than 3 m they are also aggregated. This delineation method allows to discard low vegetation, but buildings might be extracted and considered as trees. Isolated outliers do not have any impact on the results. The proposed tree delineation is illustrated in Figure 2.

2.1.2. Point-based lidar features.

Lidar-derived features require a consistent neighbourhood for their computation. For each lidar point, 3 cylindrical neighbourhoods, aligned with the vertical axis, are used (1 m, 3 m and 5 m radii, infinite height). A cylinder appears to be the most relevant environment in forested areas so as to take into account the variance of altitudes of the lidar points. Three radius values are considered so as to handle the various sizes of the trees and assuming a feature selection step will prune the initial set of attributes. Two vegetation density features, $D_1$ and $D_2$, are computed: the first one based on the number of local maxima within the neighbourhoods, and the second one related to the number of non-ground points within the neighbourhoods (ground points were previously determined by a filtering step). $D_1$ and $D_2$ are calculated as follows:

$$D_1 = \sum_{r_1 \in \{1,3,5\}} \sum_{r_2 \in \{1,3,5\}} Nt_{r_1,r_2},$$

$$D_2 = \frac{1}{3} \sum_{r \in \{1,3,5\}} \frac{Ns_r}{Ntot_r},$$

where $Nt_{r_1,r_2}$ is the number of local maxima retrieved from a $r_1$ maximum filter within the cylindrical neighbourhood of radius $r_2$. $Ns_r$ is the number of points classified as ground points within the cylindrical neighbourhood of radius $r$ and $Ntot_r$ is the total number of points within the cylindrical neighbourhood of radius $r$. Additionally, the scatter $S$ and the planarity $P$ features are computed following Weimann et al. (2015):

$$S = \frac{1}{3} \sum_{r \in \{1,3,5\}} \frac{\lambda_{3,r}}{\lambda_{1,r}},$$

$$P = \frac{1}{3} \sum_{r \in \{1,3,5\}} 2 \times (\lambda_{2,r} - \lambda_{3,r}),$$

10
where $\lambda_{1,r} \geq \lambda_{2,r} \geq \lambda_{3,r}$ are the eigenvalues of the covariance matrix within the cylindrical neighbourhood of radius $r$. They are retrieved with a standard Principal Component Analysis.

Statistical features, known to be relevant for vegetation type (mainly tree species) classification (Dalponte et al., 2014; Torabzadeh et al., 2015), are also derived. For each lidar point, the same 3 cylindrical neighbourhoods are used. Two basic information from the lidar data, namely height and intensity, are used to derive statistical features. A statistical feature $f_d$, derived from an original feature $f_o$, (height or intensity) is computed as follows:

$$f_d = \frac{1}{3} \sum_{r \in \{1,3,5\}} f_s(p_r f_o),$$

(5)

where $f_s$ is a statistical function (minimum; maximum; mean; median; standard deviation; median absolute deviation from median (medADmed); mean absolute deviation from median (meanADmed); skewness; kurtosis; 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th and 95th percentiles), and $p_r f_o$ a vector containing the sorted values of the original feature $f_o$ within the cylindrical neighbourhoods of radius $r$. All the statistical functions are used for the height. Only the mean is used for the intensity: it is hard to know how well the sensor is calibrated and a suitable correction of intensity values within tree canopies has not yet been proposed.

24 features are extracted during this step; 2 related to vegetation density, 2 related to the 3D local distribution of the point clouds (planarity and scatter), and 20 statistical features.

### 2.1.3. Pixel-based multispectral features.

The original 4 spectral bands of the image are kept and considered as multispectral features. The Normalized Difference Vegetation Index (NDVI), (Tucker, 1979), the Difference Vegetation Index (DVI), (Bacour et al., 2006) and the Ratio Vegetation Index (RVI) (Jordan, 1969) are computed as they are relevant vegetation indices. Indeed, they can provide more information about the species than the original bands alone (Zargar et al., 2011). As the point-based lidar features, statistical features are also derived from each band and each vegetation index according to Equation 5 (3 circular neighbourhoods of 1 m, 3 m and 5 m radii). Other statistical functions are used (minimum; maximum; mean; median; standard deviation; mean absolute deviation from median (meanADmed); mean absolute deviation from mean...
(meanADmean); median absolute deviation from median (medADmed); median absolute deviation from mean (medADmean). Finally, the pixel-based multispectral feature set is composed of 70 attributes.

2.1.4. Pixel-based lidar features.

The lidar features are rasterized at the same resolution of the multispectral image using a pit-free method proposed in (Khosravipour et al., 2014). This rasterization method is interesting because it produces smooth images that will lead to better results for classification and regularisation (Li et al., 2013). Such data fusion process at the feature level is valid since both datasets have approximately the same spatial resolution. The Canopy Height Model (CHM) is also computed using this method, at the same spatial resolution using an existing 1 m Digital Terrain Model provided with the filtered point cloud (Ferraz et al., 2016). The CHM is very important as it allows to derive the height above the ground and is known as a very discriminative feature for classification (Mallet et al., 2011; Weinmann, 2016).

2.1.5. Object-based feature map.

The pixel-based multispectral and lidar maps are merged so as to obtain a pixel-based feature map. Then, an object-based feature map is created using the individual trees and the pixel-based feature map. The value \( v_t \) of a pixel belonging to a tree \( t \) in the object-based feature map is computed as follows:

\[
v_t = \frac{1}{N_t} \sum_{p \in t} v_p,
\]

where \( N_t \) is the number of pixels in tree \( t \) , and \( v_p \) is the value of the pixel \( p \). If a pixel does not belong to a tree, it keeps the value of the pixel-based feature map. Here, only the mean value of the pixels within the tree is envisaged but one can also consider other statistical values (minimum, maximum, percentiles etc.).

Other morphological features could also be directly derived from the lidar cloud point at the object-level. For instance, an alpha-shape can be performed on the individual trees (Vauhkonen et al., 2010) and a penetration feature can be derived as it can help to classify vegetation type (mainly tree species). However, low point densities (1-5 points/m²) compatible with large-scale lidar surveys are not sufficient in order to derive a significant penetration indicator.
2.2. Classification

The classification is performed using a supervised classifier, in order to discriminate the vegetation type (mainly tree species) provided by an existing forest land-cover database. The classifier used in this study is the Random Forest (RF), implemented in OpenCV (Bradski and Kaehler, 2008), as it has been shown relevant in the literature (Belgiu and Drăguţ, 2016) and in a previous study compared to SVM (Dechesne et al., 2016), since it provide similar results while being faster. The outputs of the classification are (i) label map and (ii) probability map (posterior class probabilities for each pixel/object). This probability map is the main input for the subsequent regularisation step.

In order to reduce the computation times, a feature selection is carried out to identify an "optimal" feature subset (Section 2.2.1). Additionally, a strategy is proposed in order to select the most suitable training pixels for an existing land-cover forest maps, subsequently improving the classification accuracy (Section 2.2.2).

2.2.1. Feature selection

Due to the high number of features involved, an automatic Feature Selection (FS) has been integrated. This selection is composed of two steps: the choice of the number of features to select and the feature selection itself. Indeed, the choice of the number of features is very important because it enables to greatly decrease the computation times.

The Sequential Forward Floating Search (SFFS) (Pudil et al., 1994) algorithm is used for both steps. The SFFS algorithm has two main advantages: (i) it can be used with many classification score (in this study, the Kappa coefficient), (ii) it enables to access to the evolution of the classification score/accuracy according to the number of selected features. The accuracy of the classification is assessed through the Kappa coefficient of the RF classifier. The SFFS algorithm selects \( p \) features by maximizing FS score criterion (the Kappa coefficient). In order to retrieve the optimal number of features, the SFFS algorithm is performed \( n \) times on different training sets with \( p \) equal to the total number of features (95). The classification accuracy is conserved for each selection of \( s \) features \( (s \in [1, p]) \) and averaged over the \( n \) iterations. The number of optimal features \( n_{\text{opt}} \) corresponds to the size of the selection of \( s \) features having the maximal mean accuracy. In order to reduce the computation times, the optimal number of features was computed for a single 1 km\(^2\) area and used for all the areas of interest.
The feature selection is then carried out for each area of interest (one selection for each area) with $p = n_{\text{opt}}$. The selected features are used for both the classification and the energy minimisation framework. The feature selection could be carried out only once on a single area or on multiple areas in order to reduce the computation times.

2.2.2. Training set design

Using an existing forest land-cover (LC) database for training a model is not straightforward ([Gressin et al., 2013], [Radoux et al., 2014], [Maas et al., 2016]). First, locally it can suffer from a lack of information (not all the classes of interest are present). Secondly, this database may also be semantically and, more frequently, geometrically incorrect (see Figure 3): changes may have happened (forest cut or grow) and the geodatabase may have been generalized, resulting in sharp polygon vertices that do not exactly correspond to the class borders. Thirdly, in many forest LC databases, polygons of a given vegetation type (mainly tree species) may contain other vegetation type (mainly tree species) in a small proportion. Two strategies are employed to overcome these problems:

Figure 3. Standard errors in forest land-cover geodatabases. The polygon borders (in green) are superimposed onto the infra-red coloured orthoimage ©IGN. 1: Forest changes: bare soil labelled as forest – 2: deciduous trees not considered as forested areas – 3: deciduous trees labelled as fir/spruce.
Firstly, in order to increase the knowledge on existing class labels, the model could be trained on a larger area. The size of the training area has been chosen arbitrarily. However, when it is too large, we observe that the quality of the classification decreases. The optimal choice of the training area has not been investigated yet. The model has therefore been trained on a larger area than the ones of interest. The areas selected for the training are the ones maximizing the number of classes within a 5 km search zone.

Secondly, in order to correct the potential errors of the LC database, a k-means clustering has been therefore performed on each of the labels in the training area. We assume that erroneous pixels are present in a small proportion and that therefore the main cluster corresponds to the class of interest. Let $p_{i-c,t}$ be the ith pixel of the vegetation type (mainly tree species) $t$ in the cluster $c$ of the k-means. The pixels $P_t$ used to train the model for the vegetation type (mainly tree species) $t$ correspond to the set:

$$P_t = \{p_{i-c,t} \mid c = \arg\max_{c \in [1,k]} \text{Card}(\cup_i p_{i-c,t})\}. \quad (7)$$

That is to say, only samples belonging to the main k-mean cluster among training pixels for one class are kept in the training dataset.

In practice, $k = 3$: the main cluster corresponds to the label of interest whereas the two other ones correspond to the ground and minority vegetation type (mainly tree species) within the polygons. 1000 samples per class are then randomly selected in order to design the final training set.

### 2.3. Regularisation

The classification may not be sufficient so as to obtain homogeneous areas with smooth borders. Therefore, in order to fit to the stand model, classification regularisation at the pixel level, through an energy minimisation framework, appears to be a reliable solution. Additionally, it makes it possible to insert additional constraints for flexible forest stand delineation. The proposed energy model relies on a graphical model, that is to say the problem is modelled by a probabilistic graph taking into account the posterior class probabilities $P$ and the normalized (i.e., transform it to have zero mean and unit variance) pixel-based feature map $A$. It consists of a fit-to-data term and a (pairwise) regularisation term. For an image $I$ and a classification $C$, the energy $E$ is formulated as follows:

$$E(I, C) = \sum_{u \in I} E_{\text{data}}(C(u)) + \gamma \sum_{u, v \in N} E_{\text{pairwise}}(C(u), C(v)), \quad (8)$$
with:

\[
\gamma \in [0, \infty]; \\
E_{\text{data}}(C(u)) = f(P(C(u))); \\
E_{\text{pairwise}}(C(u) = C(v)) = 0; \\
E_{\text{pairwise}}(C(u) \neq C(v)) = V(u, v).
\]

\(\mathcal{N}\) is the 8-connectivity neighbourhood, and \(P(C(u))\) is the probability (according to the RF classification) that the pixel \(u\) belongs to the class \(C\). \(\gamma\) is the smoothing parameter that allows to control the degree of homogeneity of the segments (see Section 2.3.3).

\(E_{\text{data}}\) is related to the classification process. If \(E_{\text{data}}(C(u))\) is small, the pixel \(u\) has a high probability to belong to the class \(C\).

\(E_{\text{pairwise}}\) is a contrast sensitive measure corresponding to the difference between the features of the pixel \(u\) and the features of its 8 neighbours. If \(E_{\text{pairwise}}\) is high, it means the values of the features of the pixel \(u\) are different from the one of its neighbour. The total energy expresses how good the pixels are classified and how close their features are from the ones of their neighbours.

Several other Conditional Random Fields (CRF) model could be envisaged for the expression of the energy (Schindler, 2012; Volpi and Ferrari, 2015; Tuia et al., 2016), \(E_{\text{pairwise}}\) could then be expressed relatively to more pixels within the image instead of only the 8-connectivity neighbourhood. They were not adopted here.

The fit-to-data term is presented in Section 2.3.1. The regularisation term is introduced in Section 2.3.2. Finally, the energy minimisation method is detailed in Section 2.3.3.

2.3.1. Fit-to-data term.

The function \(f\) related to \(E_{\text{data}}\) (see Equation 8) is the following:

\[
f(x) = 1 - x, \quad \text{with } x \in [0, 1].
\]

(9)

This function enables to control the importance to give to the raw classification step. A simple linear function; when the posterior class probability is close to zero, the energy is maximum. Conversely, when it is close to one, the energy is minimum. This function allows to keep the energy in \([0, 1]\)
and then simplifies the tuning of the $\gamma$ parameter. Additional investigations could be envisaged on the data term and especially in the choice of the related function.

2.3.2. Regularisation term.

The regularisation term is a contrast sensitive measure that controls the value of the energy according to the value of the features within the 8-connexity neighbourhood. Two pixels with different class labels but with close feature values are more likely to belong to the same class than two pixels of different classes and different feature values. The value of the energy should be close to 0 when the feature values are rather similar and increase when they are different. The pairwise energy $V$ is expressed as follows:

$$V(u, v) = \frac{1}{n_{opt}} \sum_{i=1}^{n_{opt}} \left(1 - \exp\left(-\lambda_i \|A_i(u) - A_i(v)\|_2\right)\right).$$

(10)

$\forall i \in [1, n_{opt}]$, $\lambda_i \in [0, \infty[$, where $i$ is related to the index of a feature. $A_i(u)$ is the value of the $i^{th}$ at pixel $u$, $\|\cdot\|_2$ is the $\ell_2$ norm. $\lambda = [\lambda_1, \lambda_2, ..., \lambda_{opt}]$ is a vector of length equal to the number of features ($n_{opt}$). This vector assigns different weights to the different features. If $\lambda_i = 0$, the feature will not be taken into account in the regularisation process. When $\lambda_i$ is important, a small difference in the values of the feature will lead to an increase of the energy. Thus, feature $i$ would have a greater impact in the regularisation. As the features are of different types (height, reflectance, density, etc.), it is therefore important to have each term of the sum lying in $[0, 1]$ for each feature, even if they initially do not have the same range. In this method, $\lambda_i$ was set to 1 for all the $n_{opt}$ features that had been selected by the SFFS algorithm and $\lambda_i$ was set to 0 for the other features.

2.3.3. Energy minimisation.

The energy minimisation is performed using graph cut methods. The graph cut algorithm employed is the quadratic pseudo-boolean optimization (QPBO) with $\alpha$-expansion moves. The QPBO is a popular and efficient graph-cut method as it efficiently solves energy minimisation problems by constructing a graph and computing the min-cut (Kolmogorov and Rother, 2007). The $\alpha$-expansion is a way to deal with the multi-class problems (Kolmogorov and Zabih, 2004).

Graph-cut based methods have been used to solve many remote sensing problems (Chehata et al., 2009; Golovinskiy and Funkhouser, 2009; Reitberger, 2017).
et al. 2009; Yang et al. 2016), in particular for segmentation purpose. Several max-flow/min-cut techniques exist to solve exactly the binary case or to solve approximately the multi-label case e.g., with $\alpha$-expansion moves. Such energy minimization techniques require the energy to be sub-modular. Conversely, QPBO technique is invariant to label orderings and is therefore tailored to solve minimization problems without limiting ourselves to sub-modular energies.

For scalability purpose, the graph is computed and solved on a $N \times N$ floating window ($N = 1400$ in our experiments) and only the pixels in the $l \times l$ center window ($l = 1000$ in our experiments) are conserved. The process is iterated until the whole area is processed. Experiments have shown that the choice of $N$ and $l$ has no impact on the results ($N$ must be greater than $l$), it only aims to reduce the computational load by solving a smaller graph.

When $\gamma = 0$, the regularisation term has no effect in the energy formulation; the most probable class is attributed to the pixel, leading to the same result as the classification output. When $\gamma \neq 0$, the resulting label map becomes more homogeneous, and the borders of the segments/stands are smoother. However, if $\gamma$ is too high, the small areas are bound to be merged into larger areas, removing a part of the useful information provided by the classification step.

3. Data

3.1. Areas of interest

The test area is a mountainous forest in the East of France (Vosges). It is of high interest for forest studies due to the important vegetation type (mainly tree species) diversity in this area. 5 areas composed of a wide range of vegetation type (mainly tree species) are selected (see Figure 4 and Table 1). Area 1 is a large area ($25 \text{ km}^2$), chosen in order to evaluate the scalability of the method. Areas 2, 3 and 4 ($1 \text{ km}^2$ each) exhibit a large number of species (4-5 species per area) in order to test the robustness of the method to a large number of classes. Finally, Area 5 is selected since it contains errors visually identifiable in the forest land-cover database (see Figure 3 for more details). It allows to assess the reliability of the proposed training set selection procedure.
(a) Area 1 (25 km$^2$).

(b) Area 2 (1 km$^2$).  (c) Area 3 (1 km$^2$).  (d) Area 4 (1 km$^2$).  (e) Area 5 (1 km$^2$).

Figure 4. Orthoimages of the areas of interest ©IGN.
<table>
<thead>
<tr>
<th>Vegetation type</th>
<th>Scientific name</th>
<th>Areas of presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beech</td>
<td>Fagus sylvatica</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>Chestnut</td>
<td>Castanea sativa</td>
<td>4</td>
</tr>
<tr>
<td>Robinia</td>
<td>Robinia pseudoacacia</td>
<td>4</td>
</tr>
<tr>
<td>Deciduous oaks</td>
<td>Deciduous quercus</td>
<td>1, 2, 4</td>
</tr>
<tr>
<td>Scots pine</td>
<td>Pinus sylvestris</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>Douglas fir</td>
<td>Pseudotsuga menziesii</td>
<td>2, 3, 4</td>
</tr>
<tr>
<td>Fir or Spruce</td>
<td>Abies or picea</td>
<td>1, 2, 3, 4, 5</td>
</tr>
<tr>
<td>Woody heathland</td>
<td></td>
<td>1, 5</td>
</tr>
<tr>
<td>Herbaceous formation</td>
<td></td>
<td>1, 5</td>
</tr>
</tbody>
</table>

Table 1. List of the vegetation type (mainly tree species) and their areas of presence.

3.2. Data specifications

The airborne multispectral images were captured by the IGN digital cameras \[\text{[Souchon et al. 2012]}\]. They have 4 bands: 430-550 nm (blue), 490-610 nm (green), 600-720 nm (red) and 750-950 nm (near infra-red) at 0.5 m ground sample distance (spatial resolution).

The airborne lidar data were collected using an Optech 3100EA device. The footprint was 0.8 m in order to increase the probability to reach the ground. The point density for all echoes ranges from 2 to 4 points/m². Our multispectral and lidar data fit with the standards used in many countries for large-scale operational forest mapping purposes \[\text{[White et al. 2016]}\]. Data were acquired under leaf-on conditions, in May and June 2011 for the multispectral images and the lidar data, respectively. Operating under leaf-off conditions would be more beneficial for tree segmentation but to the detriment of tree species diversity. The registration between airborne lidar point clouds and VHR multispectral images was performed by IGN itself using ground control points. This is a standard procedure in the French mapping agency and it has results comparable with standard professional solutions.

The forest LC geospatial database is composed of 2D polygons delineated by photo-interpreters. It is the French national LC datum for forests, freely available for end-users.\[\text{1}\] It is used in order both to train the classifier and to evaluate the results. Only the polygons containing at least 75% of a given

---

\[1\]http://inventaire-forestier.ign.fr/spip/?rubrique67

20
species are used for the classification task (this is the threshold which defines when a stand can be assigned to a unique vegetation type). The polygons of natural bare soils (woody heathland and herbaceous formation) are also used for the classification. As this study is only based on species, the ground truth used will therefore not cover the entire area (excluding the test areas containing stands of mixed species).

4. Experiments and evaluations

Various tests have been carried out in order to validate the method. First, the feature selection, crucial step for the classification and the regularisation, is investigated (Section 4.1). The classification process and its impact on the final stand segmentation is then presented (Section 4.2). The tuning of the regularisation procedure is discussed in Section 4.3. The advantages of the fusion of ALS data and multispectral images are presented in Section 4.4. Finally, the computational effort of the proposed method is presented in Section 4.5.

4.1. Feature selection

The SFFS algorithm allows to determine the optimal number of features, and select them in a forward way (see Section 2.2.1). In order to estimate the number of optimal features, the SFFS algorithm was performed 50 times on Area 4, which is the most diverse in terms of species. This step was highly time consuming (about 50 hours) and was thus performed just once. The number of optimal features to select was found to be 20. This number was conserved for the other areas and the results show that the transfer has no impact on the final accuracy.

Once the optimal number of features was determined, the feature selection was performed 40 times over all the test areas in order to retrieve the most relevant features. The retained attributes are presented in Figures 5, 6 and 7. On average, 61% of the selected features are derived from the spectral information and 39% from the lidar information. This shows the complementarity of both remote sensing data.

For the spectral information, the features derived from the original band set are more relevant than the vegetation indices: the near-infrared derived features represent 18% of the spectral selected features, 16% for the red and the green, 15% for the blue and the DVI, only 11% for the NDVI and 10% for the RVI. The most relevant statistical feature for the spectral information
is the minimum (17% of the spectral selection). The maximum (12%), the median (11%), the mean (11%) and the standard deviation (10%) are also relevant. The other statistics are selected at less than 9% each.

For the lidar information, the most relevant feature is surprisingly the intensity, selected in each of the 40 selections (12% of the lidar selection, 5% of the total selection). The standard deviation (8% of the lidar selection), the maximum (7%) and the densities (5% and 6%) are also relevant. The other lidar derived features count for less than 4% each.

### Spectral-Derived Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Red Band</th>
<th>Green Band</th>
<th>Blue Band</th>
<th>Near-Infrared Band</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Selections</strong></td>
<td><img src="image1" alt="Red-band derived features" /></td>
<td><img src="image2" alt="Green-band derived features" /></td>
<td><img src="image3" alt="Blue-band derived features" /></td>
<td><img src="image4" alt="Near-infrared-band derived features" /></td>
</tr>
</tbody>
</table>

Figure 5. Spectral-derived features selected over 40 selections of 20 optimal features.

### 4.2. Classification

The classification results and their impact on the final segmentation are illustrated in Figure 8 with Area 3. The accuracy is obtained by comparing each pixel of the forest LC database to the classification results. It clearly
Figure 6. Vegetation index-derived features selected over 40 trials of 20 optimal features.
Figure 7. Lidar-derived features selected over 40 trials of 20 optimal features.
appears that the pixel-based approach leads to noisy label maps, even if correct results are obtained (79\% of overall accuracy, see Figure 8b). Conversely, even if the tree extraction is rough, the object-based feature map leads to more spatially consistent labels (84\% of overall accuracy, see Figure 8c). This difference is directly reflected on the final segmentation. Indeed, the regularisation based on the pixel-based classification has 90\% (Figure 8e) of correct match with the LC database, while the regularisation based on the object-based classification has a matching score of 97\% (Figure 8f). These results confirm the relevance of the tree-level, even coarsely defined, for the classification. Theses results have also been observed for the other areas.

Figure 8. Classification and segmentation results for Area 3; comparison between the pixel- and object-based approaches. Colour codes: \textcolor{red}{beech}, \textcolor{blue}{Scots pine}, \textcolor{green}{Douglas fir}, \textcolor{cyan}{fir or spruce}, \textcolor{black}{no data}.

The effect of the selection of the training pixels is presented in Figure 9.
for Area 5. When no selection is operated (random choice), the bare soil in the forest polygon is interpreted as forest, resulting in a highly aberrant segmentation. The use of the k-means algorithm on the initial LC polygons allows to overcome this problem. However, for this area, the class that are represented (i.e., deciduous trees) but not labelled in a polygon might be aggregated to the wrong class.

Figure 9. Classification and segmentation results; the effect of a robust selection of training pixels for Area 5. Colour codes: ⬤ fir or spruce, ⬤ woody heathland, ⬤ herbaceous formation, ⬤ no data.

4.3. Regularisation
The effect of the smoothing parameter $\gamma$ is presented in Figures 10 and 11.
When $\gamma$ is too low (e.g., $< 30$), some small irrelevant segments might remain in the final segmentation, leading to a low correct matching rate with the LC database. While inefficient for change detection purposes, this over-segmentation may be useful for statistical forest inventory and cartography, as it provides small homogeneous areas within large stands. When $\gamma$ lies in $[40, 70]$, the segmentation accuracy is high: each class is correctly labelled at more than 95%. However, having a high $\gamma$ ($> 90$) slightly decreases the segmentation accuracy but also greatly reduces the accuracy of under represented classes (e.g., deciduous oak is totally merged in fir or spruce). These results have also been observed on the other areas, leading to the conclusion that $\gamma$ values should lie in $[40, 70]$.

![Figure 10. Effect of the smoothing parameter $\gamma$ on the segmentation accuracy for each class of Area 2 (see Figure 11). Colour codes: beech, deciduous oaks, Scots pine, Douglas fir, fir or spruce, black: overall accuracy.](image)

The final results are presented in Table 2 and Figures 12, 13, and 14. The overall accuracy shows first that the method gives satisfactory results in terms of vegetation type (mainly tree species) discrimination, in the range
Figure 11. Effect of the smoothing parameter $\gamma$ on the segmentation for Area 2. Colour codes: Ⓡbeech, Ⓡdeciduous oaks, ⓇScots pine, ⓇDouglas fir, Ⓡfir or spruce, Ⓡno data.
of existing papers of the literature for the same number of species (Leckie et al., 2003; Leppänen et al., 2008). Secondly, results are improved up to 15\% when smoothing the results. Despite very high accuracies, the results have to be considered with care. They are compared with a forest LC database, which has its own flaws and does not cover 100\% of the study areas. Visual analysis of the five areas allows us to say that the accuracy scores can in fact be taken in a range of ±5\%. For instance, Area 5 exhibits a lower segmentation accuracy because of a larger proportion of errors in the forest LC database (bare soil labelled as forest).

The results are very similar for Areas 2, 3 and 4. When two classes do not have a common boundary, the method performs better. Indeed, if the boundary is not exactly retrieved, the results will not be impacted (see Area 2). However, results are still satisfactory when two classes have a common boundary (as for Areas 3 and 4). Area 1 exhibits slightly worse results than the other areas. The classification performance tends to be more limited at a larger scale, directly impacting the segmentation results. However, the strategy of computing the graph on a $N \times N$ window and keeping the pixel in $l \times l$ center window works well since contiguous regions with high variability in size can be extracted (see Figure 12). Stands of a pure species may exhibit a great variability both in age and height, explaining why a large-scale learning procedure is bound to fail for forested areas. More homogeneous stands could be retrieved afterwards, based on height criterion for instance. We foresee to develop such approach in the near future but no ground truth exists to validate such a proposed approach. From a visual point of view, the results appear to be correct (the boundaries are consistent with the orthoimage). Training at larger scale does not help to improve the results in this area. It may be due to the steep slopes of the terrain. Trees may exhibit different orientations with respect to the sun for the same specie. Illumination changes may create confusion in the spectral features and subsequently perturb the learning step. Additionally, in a LC database polygon, various ages of a species may exist, resulting also in varying appearances that are not yet taken into account in our method: the selection of training pixels might remove this variability, leading to a poor accuracy for the class of interest. The accuracy of each species for Area 1 is presented in Table 3. In this area, the well represented classes are the ones that have a correct proportion (> 5\%) in the forest LC and the under represented classes are the ones that have a low proportion (< 5\%) in the forest LC. For the well represented classes (beech, fir or spruce and woody hearthland) the accuracy increases whereas
for the other classes, which are under represented (*herbaceous formation*, *deciduous oaks*, and *Douglas fir*), the accuracy decreases. Indeed, a small error on an under represented class will lead to an important decrease in the class accuracy whereas it will only have a small effect on a well represented class. This quantitative assessment will be refined in our future works with the help of forest experts.

<table>
<thead>
<tr>
<th># vegetation types</th>
<th>Classification accuracy</th>
<th>Segmentation accuracy</th>
<th>γ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area 1</td>
<td>6</td>
<td>72.61%</td>
<td>86.67%</td>
</tr>
<tr>
<td>Area 2</td>
<td>5</td>
<td>76.87%</td>
<td>99.12%</td>
</tr>
<tr>
<td>Area 3</td>
<td>4</td>
<td>84.20%</td>
<td>97.33%</td>
</tr>
<tr>
<td>Area 4</td>
<td>4</td>
<td>79.85%</td>
<td>94.40%</td>
</tr>
<tr>
<td>Area 5</td>
<td>3</td>
<td>82.97%</td>
<td>84.51%</td>
</tr>
</tbody>
</table>

Table 2. Stand segmentation accuracy obtained before and after regularisation, with respect to the ground truth provided by the forest LC database.

<table>
<thead>
<tr>
<th>Vegetation type</th>
<th>Classification accuracy</th>
<th>Segmentation accuracy</th>
<th>Proportion in the LC database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beech</td>
<td>70.15%</td>
<td>74.92%</td>
<td>19.36%</td>
</tr>
<tr>
<td>Fir or spruce</td>
<td>76.90%</td>
<td>95.96%</td>
<td>68.03%</td>
</tr>
<tr>
<td>Herbaceous formation</td>
<td>67.56%</td>
<td>60.33%</td>
<td>0.75%</td>
</tr>
<tr>
<td>Deciduous oaks</td>
<td>60.06%</td>
<td>58.44%</td>
<td>0.77%</td>
</tr>
<tr>
<td>Woody heathland</td>
<td>55.09%</td>
<td>65.30%</td>
<td>7.78%</td>
</tr>
<tr>
<td>Douglas fir</td>
<td>44.02%</td>
<td>26.76%</td>
<td>3.30%</td>
</tr>
</tbody>
</table>

Table 3. Accuracy obtained before and after regularisation for each class of Area 1, with respect to the ground truth provided by the forest LC database.

4.4. Advantages of multi-modal remote sensing

The method has been tested on Area 4 using a single data source (only lidar or multispectral), in order to assess the complementary of both kinds of sensors. For the lidar data, since only 25 features are available, all were used in the classification and regularisation processes. Because of the larger
Figure 12. Results for Area 1. Colour codes: ⬜ beech, ⬜️ deciduous oaks, ⬜️ Scots pine, ⬜️ fir or spruce, ⬜️ woody heathland, ⬜️ herbaceous formation, ⬜️ no data.
Figure 13. Segmentation results for areas 2 and 3. Colour codes: 
- beech, 
- chestnut, 
- robinia, 
- deciduous oaks, 
- Scots pine, 
- Douglas fir, 
- fir or spruce, 
- woody heathland, 
- herbaceous formation, 
- no data.
Figure 14. Segmentation results for areas 4 and 5. Colour codes: 
- beech, 
- chestnut, 
- robinia, 
- deciduous oaks, 
- Scots pine, 
- Douglas fir, 
- fir or spruce, 
- woody heathland, 
- herbaceous formation, 
- no data
number of multispectral features (70), a selection of 20 features was carried out. The selected features were both used for the classification and regularisation steps. In terms of performance, lidar data provides worse results: the classification has an overall accuracy of 52%. The regularisation step does not significantly improve the results, with an overall accuracy of \( \sim 60\% \). The multispectral data produce better results; the classification has an overall accuracy of 57% and the regularisation tends to \( \sim 75\% \) in the range of suitable \( \gamma \) parameters. Multispectral images are particularly beneficial for the regularization step, with respect to lidar point clouds (\( +5\% \) after classification \( \rightarrow +10\text{-}15\% \) after segmentation). Best results are obtained after fusing the two data sources, with a significant impact in terms of accuracy (\( +20\% \)). The results are presented in Figure 15 and Table 4.

<table>
<thead>
<tr>
<th></th>
<th>Classification accuracy</th>
<th>Segmentation accuracy</th>
<th>( \gamma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALS</td>
<td>52.18%</td>
<td>60.44%</td>
<td>30</td>
</tr>
<tr>
<td>ALS</td>
<td>52.18%</td>
<td>61.31%</td>
<td>50</td>
</tr>
<tr>
<td>Multispectral</td>
<td>57.75%</td>
<td>74.47%</td>
<td>30</td>
</tr>
<tr>
<td>Multispectral</td>
<td>57.75%</td>
<td>74.91%</td>
<td>50</td>
</tr>
<tr>
<td>ALS + Multispectral</td>
<td><strong>79.85%</strong></td>
<td><strong>94.40%</strong></td>
<td>50</td>
</tr>
</tbody>
</table>

Table 4. Stand segmentation accuracy without and with data fusion, both before and after the regularisation process. The ground truth is provided by the forest LC database for the Area 4.

4.5. Computation times

They are presented in Table 5 where they are averaged over the 5 areas for an equivalent of a 1 km\(^2\) area. Firstly, the tree delineation is time consuming although coarse results are targeted. The computation of the features is fast, given the number of features. The feature selection has reasonable computation time. The classification is very fast, similarly to the regularisation process. Yet, no investigation have been carried out in order to improve the computation times.

5. Conclusion

A 3-step method for forest stand delineation according to vegetation type (mainly tree species) was proposed. The fusion of ALS data and multispectral
Figure 15. Stand segmentation results with different source of data before and after regularisation the Area 4. Colour codes: ● chestnut, ● robinia, ● deciduous oaks, ● fir or spruce, ● no data.
images produces very satisfactory results since both remote sensing modalities provide complementary observations. The final segments have a good matching score with the stands delineated by human operators. The method relies on the computation of lidar and multispectral features at different levels (pixel and tree) for a supervised classification of vegetation type (mainly tree species). Good discrimination scores are already obtained with standard features and classifier, which is a strong basis for an even more accurate delineation. This delineation is then achieved thanks to a regularisation. It relies on an energetical model formulated according to both classification posterior probabilities results and feature values. It enables to obtain homogeneous areas in terms of species with smooth borders. It is highly helpful in order to control the level of details required for the stand delineation, which depends on forest inventory or land-cover database specifications. However, at large scale, the performance of the method decreases.

Since the aim is to delineate forest stands according to vegetation type (mainly tree species), the introduction of hyperspectral images might be interesting so as to obtain more information about the species, and especially overcome the problem of species variability. Some other vegetation indices can also be derived from hyperspectral data. Similarly, the use of higher density ALS data ( ~ 10 pts/m²) might also improve the results of the method; trees could be extracted more precisely with an improved tree delineation technique and new structural features could subsequently be derived at the tree level (Ørka et al., 2009).

For land-cover mapping, the proposed framework will be modified in order

<table>
<thead>
<tr>
<th></th>
<th>Computation times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree extraction</td>
<td>~ 1 h 40 min</td>
</tr>
<tr>
<td>Lidar features</td>
<td>~ 1 h</td>
</tr>
<tr>
<td>Multispectral features</td>
<td>~ 1 h</td>
</tr>
<tr>
<td>Object-based feature map</td>
<td>~ 5 m</td>
</tr>
<tr>
<td>Feature selection</td>
<td>~ 45 min</td>
</tr>
<tr>
<td>Classification</td>
<td>~ 1 min</td>
</tr>
<tr>
<td>Regularisation</td>
<td>~ 10 min</td>
</tr>
<tr>
<td><strong>Full algorithm</strong></td>
<td>~ 4 h 45 min</td>
</tr>
</tbody>
</table>

Table 5. Computation times of the different steps of the algorithm (per km²).
to be able to retrieve different levels of standard forest databases in a hierarchical way (vegetation/non-vegetation → deciduous/coniferous → tree species), still with the same inputs (trees and species). Finally, we aim to deal with mixed forests (i.e., forest stands without dominant tree species) and open forests (i.e., canopy cover between 10% and 40%) in order to be able to retrieve all conceivable forest labels.

References


