# On the use of binary partition trees for the tree crown segmentation of tropical rainforest hyperspectral images

G. Tochon[a,b], J.B. Féret[b,c], S. Valero[c], R.E. Martin[b], D.E. Knapp[b], P. Salembier[d], J. Chanussot[a,e], G.P. Asner[b]

[a]*GIPSA-Lab, 11 rue des Mathématiques, 38400 Saint Martin d'Hères, France.*
[b]*Department of Global Ecology, Carnegie Institution for Science, 260 Panama Street, Stanford, CA 94305, USA.*
[c]*CESBIO, 18 avenue Edouard Belin, 31400 Toulouse, France.*
[d]*Technical University of Catalonia (UPC), Jordi Girona 1-3, edifici D5, 08034 Barcelona, Spain.*
[e]*Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavík, Iceland.*

**Abstract**

The segmentation of remotely sensed images acquired over tropical forests is of great interest for numerous ecological applications, such as forest inventories or conservation and management of ecosystems, for which species classification techniques and estimation of the number of individuals are highly valuable inputs. In this paper, we propose a method for hyperspectral image segmentation, based on the Binary Partition Tree (BPT) algorithm, and we apply it to two sites located in Hawaiian and Panamean tropical rainforests. Different strategies combining spatial and spectral dimensionality reduction are compared prior to the construction of the BPT. Various superpixel generation methods including watershed transformation and mean shift clustering are applied to decrease spatial dimensionality and provide an initial partition map. Principal component analysis is performed to reduce

*Email addresses:* guillaume.tochon@gipsa-lab.grenoble-inp.fr (G. Tochon), feretjb@cesbio.cnes.fr (J.B. Féret)

the spectral dimensionality and different combinations of principal components are compared. A non-parametric region model based on histograms, combined with the diffusion distance to merge regions, is used to build the BPT. An adapted pruning strategy based on the size discontinuity of the merging regions is proposed and compared with an already existing pruning strategy. Finally, a set of criteria to assess the quality of the tree segmentation is introduced. The proposed method correctly segmented up to 68% of the tree crowns and produced reasonable patterns of the segmented landscapes.
*Keywords:* Binary partition tree, Carnegie Airborne Observatory, hyperspectral imagery, segmentation, tree crown delineation, tropical forest

## 1. Introduction

There is a growing need for large-scale assessment of biodiversity and species richness in ecosystems, as a means to improve forest conservation and management decisions. Tropical rainforest ecosystems are of critical interest since they are hotspots of biodiversity, greatly contributing to the worlds biotic variety while covering only a small percentage of the terrestrial surface. Moreover, they are particularly vulnerable to multiple factor pressures such as exploitation of natural resources and climate change (Asner et al., 2009; Thomas et al., 2004; Whitmore et al., 1990). In this context, information about the forest structure, the number, spatial distribution and identification of individual trees, the species richness and its evolution, and the dynamics of invasive species across landscapes are highly sought after for efficient management decisions applied to forest conservation. Related field data collection is extremely expensive, time-consuming and requires very skilled field workers. Such constraints call for supporting technologies and methods for the accurate and regular monitoring of the evolution of biological diversity over large spatial scales. Remote sensing appears as a particularly efficient tool for such applications (Rasi et al., 2013; Reiche et al., 2013). However, monitoring tropical forest ecosystems using remote sensing remains extremely challenging due to the complexity of the canopy in terms of density, structure and species richness (Papes et al., 2013; Pouteau & Stoll, 2012; Somers & Asner, 2013).

Among the various information that can be derived from remotely sensed data, individual tree crown (ITC) delineation is a particularly important product assisting in fine-scale analysis ecological processes linked to vegetation structure and gap dynamics (Phinn et al., 2008), as well as improved tree species identification (Clark et al., 2005). Indeed, region properties (texture, size, shape, radiometric variability) can be derived from each ITC delineated on an image, resulting in the combination of spatial and radiometric information. Such object-oriented approaches usually outperform traditional pixel-based methods for classification and other image processing applications such as spectral unmixing and object detection, and dramatically enriches contextual information delivered by remote sensing products. In practice, high spatial resolution ITC delineation can be useful to help monitor species of interest, such as dominant trees, rare or invasive species that are key indicators for environmental processes (Asner et al., 2008). It can also be used to detect illegal logging, as logging practices are nowadays very selective and assisted by moderate resolution satellite images to detect large scale deforestation (Asner et al., 2005).

Several segmentation methods have been developed for ITC delineation based on high spatial resolution imagery derived from various sensors, from satellite very high resolution imagery to airborne Light Detection and Ranging (LiDAR) data. However, the selection of a segmentation algorithm is critical as the performances of these methods are usually strongly ecosystem-dependent. ITCs that are typically encountered in temperate forests offer several appealing characteristics for the development of segmentation algorithms. In fact, those trees have a regular shape and elongated silhouette, and the canopy is rather sparse. Existing segmentation algorithms devoted to the segmentation of temperate forests are taking advantage of those properties. For instance, it is often assumed in forested area high resolution digital imagery that an ITC is represented by bright pixels (the top of the crown, well illuminated by the sun) surrounded by darker pixels (either shaded portions of the crown or the ground) (Wulder et al., 2000). Using a topographical analogy, an ITC can be viewed as a peak and the valleys circling around it are its physical boundaries. The valley following approach exploits this idea by encircling bright pixels with darker boundaries, and was used by Gougeon (1995); Leckie et al. (2005, 2003) for the segmentation of coniferous plantations, and by Warner et al. (1998) for deciduous forests. Also relying on the topographical representation, region growing approaches implement seeds in local maxima of the image, each seed being therefore potentially located at the top of an ITC. Regions are gradually expanded from the seeds until a stopping criterion, based on the presence of valleys, is reached. Region growing methods were validated on Australian eucalypt forests by Culvenor (2002); Whiteside & Ahmadb (2008) and on coniferous forests by Erikson (2004); Pouliot et al. (2002).

The marker-controlled watershed methods is analogous to region growing when grey tones are inverted in the topographical representation, that is, when local maxima corresponding to ITCs become local minima. Instead of expanding regions from bright values to dark ones, the watershed floods up the topographical map and creates regions corresponding to catchment basins. Markers play the same role as seeds in the region growing approach, and temper the algorithm's sensibility to noise in order to avoid over-segmentation. This approach was validated by Wang et al. (2004) for the segmentation of Canadian coniferous forests. A comparison between valley following, region growing and marker-controlled watershed methods for coniferous and deciduous tree stands is drawn by Ke & Quackenbush (2011). Template matching can also be applied when all ITCs have a regular and elongated shape. It consists of synthetically modelling the tree shapes by a collection of templates being generalized ellipses with various physically possible parameter values. Each template is cross-correlated against any potential tree position in the digital image, and the location of the highest correlations are considered to be ITC positions while the corresponding templates are assumed to be the tree shapes. Template matching was used by Olofsson (2002); Pollock (1996, 1998) for coniferous and mixed forests, and a comparison between template matching and region growing approaches applied to the delineation of Swedish spruce stands can be found in Erikson & Olofsson (2005). Finally, stochastic point process methods model the image as a realization of a marked point process of ellipses. The process, being the digital image, contains an unknown number of objects (trees), each of them being in an unknown configuration (the elliptic shape and orientation). An energy term corresponding to the fit between the model and the real image is defined, and the model is iteratively adjusted in order to decrease the energy term at each iteration. Prior knowledge about the general distribution of shapes and sizes is needed to operate the method, and those parameters can be easily derived when all trees have similar structures. Point processes were investigated by Perrin et al. (2005) for poplar plantations and by Andersen (2003) for coniferous forests.

These methods, based on strong hypotheses about crown size and shape (existence of one unique maximum for each individual and limited overlapping between individuals) show good results for high resolution digital images of temperate forests. However, they perform poorly when applied to tropical dense forest ecosystems, where tree size and shape are highly variable, and individuals usually overlap. Varekamp & Hoekman (2001) proposed a method based on Fourier parameterized deformable models for Interferometric Synthetic Aperture Radar (InSAR) data. Using the intensity, the interferometric height-coordinate and the coherence magnitude measures proper to the InSAR imaging system, they match ITCs with deformable ellipses, and applied their method to a tropical forest located in Kalimantant, Indonesia. Note that Zhou et al. (2010) also applied marked point processes to high resolution imagery and LiDAR-derived canopy height in order to detect individuals in high biomass mangroves, including only one to two canopy species. Results were encouraging; however they may not be replicable when applied to dense tropical forests given the relatively low heterogeneity of mangroves.

Over the last decade, several studies explored the potential of spectroscopic imagery for the tree species identification in dense tropical forests (Clark et al., 2005; Feret & Asner, 2013), as well as tree crown delineation (Bunting

& Lucas, 2006) in open mixed forests. The differentiation between species is based on their spectral signature, which is related to leaf chemistry and individual tree structure. Detailed spectral information may then be a valuable input to detect boundaries between neighboring trees in dense tropical forests. However, it comes with a major challenge related to the high dimensionality of the data and the need of adapted algorithms for automated tree crown segmentation. To the best of our knowledge, there is no reference study for the segmentation of tree crowns in hyperspectral images of tropical rainforests.

Image segmentation applied to dense tropical forests is an ill-posed task: a given image can often be segmented at several levels of details, due to the complex architecture of the top of the canopy. For this reason, it is better to have a consistent hierarchy of segmentations rather than a collection of minimally related segmentations. This allows the user to tune the exploration level within the hierarchy to the precise goal (Jung et al., 2014; Tarabalka et al., 2012). Mathematical tree structures are well suited for a hierarchical region-based representation of an image. In such structure, each node of the tree represents a given region in the corresponding image, and links between nodes illustrate a particular relationship between regions, such as inclusion or adjacency. Among all tree representations, the binary partition tree (BPT) has received much attention lately. Initially proposed by Garrido (2002); Salembier & Garrido (2000) for grayscale and RGB images, BPTs have then been further extended to hyperspectral imagery by Valero et al. (2013a) and are now used for classical hyperspectral remote sensing tasks such as segmentation (Valero et al., 2011a; Veganzones et al., 2014), classification (Alonso-Gonzalez et al., 2013), unmixing (Drumetz et al., 2014) and object detection (Valero et al., 2013b, 2011b) notably. The efficiency of the BPT to achieve a given task is greatly impacted by both the pre-processing applied to the image prior to the construction of the BPT and the post-processing of the BPT representation itself, called pruning.

Consequently, we propose in the following study to adapt the BPT representation to the segmentation of hyperspectral images of tropical rainforests, through an adapted pre-processing of the data and pruning of the BPT. The pre-processing stage consists of spectrally and spatially reducing the data by extracting discriminant information using Principal Component Analysis (PCA) and spatial pre-segmentation, respectively. Different configurations for the PCA reduction as well as several pre-segmentation algorithms are investigated. A novel BPT pruning strategy, dedicated to the segmentation of tree crowns is proposed and compared against an already existing pruning strategy. A method to assess the quality of the resulting segmentation is also introduced, allowing to state which is the most efficient spectral reduction configuration and pre-segmentation algorithm in a given context. The proposed method is tested on two data sets with different characteristics.

The paper is organized as follows: Section 2 introduces the data used to test our algorithm. Section 3 presents the methodology, namely the pre-processing operations, the construction and pruning of the BPT, and the method developed to assess the performance of the segmentation. The results are introduced and discussed in Section 4. Finally, some conclusions and perspectives for the application of our method are given in Section 5.

## 2. Materials

Two sites were selected to conduct this study. The first site, hereafter named Hawaii, is located at the Nanawale Forest Reserve, Hawaii (USA). The Nanawale forest is classified as lowland humid tropical forest, with an average elevation of 150 m above sea level. Mean annual precipitation and temperature are 3200 $mm.yr^{-1}$ and 23°C, respectively. The forest canopy is comprised of about 17 species, mostly invasive non-native trees, with a few native species remaining. The remote sensing data used in this study were acquired with the Carnegie Airborne Observatory (CAO) Alpha sensor package in September 2007 (Asner et al., 2007). The CAO-Alpha is equipped with a spectroscopic imager measuring up to 72 bands in the visible and near infrared (VNIR) domain, as well as a small footprint Light Detection and Ranging sensor (LiDAR) working simultaneously. This first study site corresponds to a 1980 by 1420 pixel image with 0.56 m ground sampling distance, covering an area of about 70 hectares on the ground. The spectral resolution used for this campaign results in the acquisition of 24 spectral bands of 28 nm in width and evenly spaced between 390 nm and 1044 nm. The LiDAR acquisitions were performed in discrete return mode (4 returns) and both digital elevation model (DEM) and canopy height model (CHM) coregistered with hyperspectral data were produced.

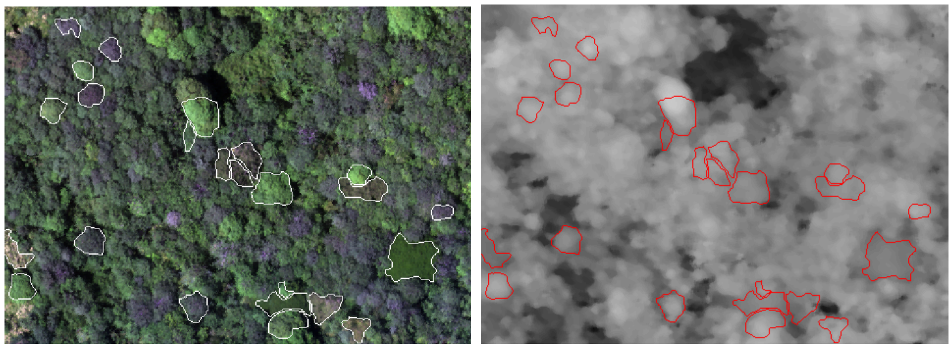
The second site, hereafter named Panama, is situated in the Panama forest. The data were collected over the Parque Nacional San Lorenzo in the Republic of Panama. The site is humid tropical forest with a mean annual precipitation of 3300 $mm.yr^{-1}$. Mean annual temperature is 26°C. The canopy is considered old growth forest populated by trees of 200-300 years of age. Canopy height ranges from about 20 m to a maximum of 45 m. The data was collected using the Carnegie Airborne Observatory Airborne Taxonomic Mapping System (CAO-AToMS) (Asner et al., 2012), launched in June 2011. The CAO-AToMS integrates three sensors in the same platform, including a new High Fidelity Visible-Shortwave Imaging Spectrometer (VSWIR) measuring the 380-2510 nm wavelength range at up to 5 nm spectral resolution, a dual-laser, waveform LiDAR system, and a high-resolution Visible-to-Near Infrared (VNIR) imaging spectrometer. The data acquired over the study site corresponds to a 600 by 600 pixels VSWIR image with a spatial resolution of 2 m and including 224 spectral bands (12 nm FWHM) evenly spaced between 378 nm and 2510 nm and co-registered DEM and CHM. 175 bands were retained from the VSWIR image after the elimination of unwanted spectral bands such as those corresponding to atmospheric water absorption.

A total of 160 ITCs for Hawaii and 100 ITCs for Panama were manually delineated by a trained operator, using the ENVI software, after visual interpretation of the hyperspectral imagery, in order to assess segmentation accuracy. Particular care was taken to include individuals of various shape, size and species. Some examples of manually delineated ITCs can be observed in Figures 1 and 2.

## 3. Proposed segmentation strategy

### *3.1. Principle of the Binary Partition Tree*

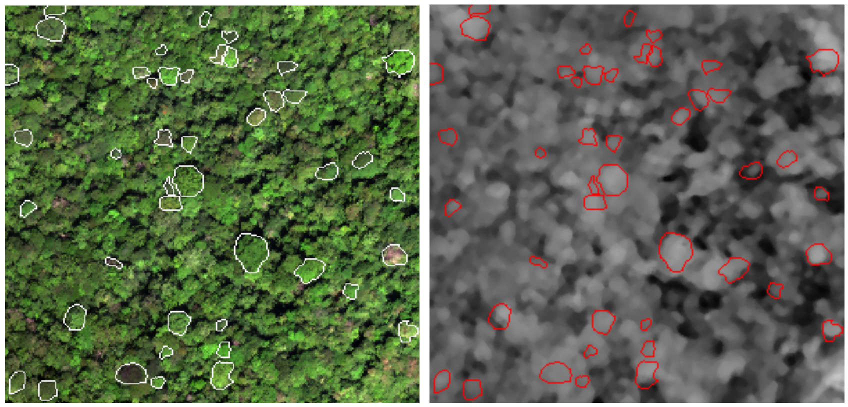
A remotely sensed image of the Earth surface is typically composed of several semantic regions of interest, such as buildings, trees, crop fields, … Those regions often follow a hierarchical organization (for instance, a building



(a) (b)

Figure 1: (a) RGB colored composition of a hyperspectral sub-image of Hawaii site (R=646 nm, G=561 nm, B=447 nm, stretched colors) with some ITCs manually delineated (in white) and (b) corresponding canopy height model derived from LiDAR with ITCs (in red).

is enclosed in a neighborhood, which is itself enclosed in a city), and the place of a particular region in a hierarchy is directly related to the scale of exploration (the scale of exploration of a building is finer than the one of a neighborhood). When analyzing an image, one has to choose a scale based on the intended level of details, and this operation is application-dependent. As a result, it can be valuable to represent the image in a task-independent hierarchy of regions, and set the exploration level in this hierarchy afterwards based on the application. The binary partition tree (BPT) is a solution to achieve such hierarchical region-based representation of an image. Starting from an initial partition of the image (corresponding to individual pixels or regions defined by a preliminary segmentation), neighboring regions are iteratively merged together until there is only one region remaining, and



(a) (b)

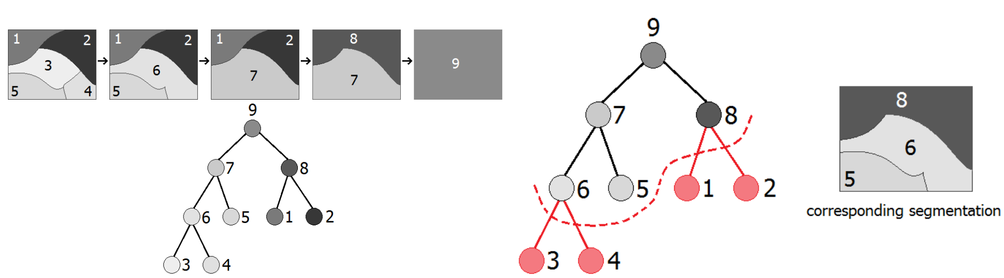
Figure 2: (a) RGB colored composition of a hyperspectral sub-image of Panama site (R=634 nm, G=549 nm, B=463 nm, stretched colors) with some ITCs manually delineated (in white) and (b) corresponding canopy height model derived from LiDAR.

those merging are stored in a tree structure. Thus, in the corresponding tree representation, the *leaf* nodes correspond to the regions in the initial partition of the image, the *root* of the tree represents the whole image support, and each node in between corresponds to the region resulting from the merging of two children regions. Following this definition, the tree structure corresponding to an initial partition of $N$ leaves contains a total of $2N-1$ nodes. Figure 3a shows the different steps of the construction of a BPT, which is determined by two notions:

- The *region model* $\mathcal{M}_{\mathcal{R}}$, which specifies how a region $\mathcal{R}$ is mathematically handled, and how to model the union of two regions. This region descriptor (for instance the mean grayscale value in figure 3a) is used to compare neighboring regions.

- The *merging criterion* $\mathcal{O}(\mathcal{R}_i, \mathcal{R}_j)$, which quantifies the similarity between neighboring regions $\mathcal{R}_i$ and $\mathcal{R}_j$ by measuring a distance between their region models. Thus, the merging criterion determines in which order the regions are merged.

The *pruning* step takes place once the BPT construction is completed. The pruning aims at cutting off branches in the BPT so the new leaves of the pruned tree achieve the most relevant segmentation regarding the application. If the construction of the tree is generic up to the definition of a region model and a merging criterion, the pruning strategy is application dependent. Therefore, the level of exploration is defined through the pruning operation, and two different pruning strategies applied on the same BPT are likely to produce different segmentations. A pruning operation is illustrated in figure 3b.




(a) (b)

Figure 3: (a) construction of a BPT, and (b) an example of pruning of it.

### *3.2. Methodology*

The proposed method is summarized by the flowchart displayed in figure 4. It is composed of 4 different steps, namely the pre-processing stage, the construction of the BPT, the pruning of the BPT and the quality evaluation of the produced segmentation map. The pre-processing stage comprises data dimensionality reduction and pre-segmentation, producing inputs for the BPT construction stage. By varying these input configurations, we study their influence on the whole segmentation and quality evaluation processes. Moreover, we introduce a BPT pruning strategy based on the evolution of the region size along branches of the BPT. We compare this new pruning strategy with an already existing one which relies on spectral graph partitioning (Alonso-Gonzalez et al., 2013; Valero, 2011). Finally, we present the metrics used for the quality assessment of segmentation maps.

### *3.3. Pre-processing step*

The construction of the BPT is computationally very intensive and may become problematic for applications on large datasets. Here, we pre-processed the data in order to reduce both spectral and spatial dimensions of the data.

#### *3.3.1. Spectral reduction*

The detailed spectral properties of an element (pixel or object) extracted from spectroscopic imagery are particularly interesting for classification purposes. However, strong correlations exist between most of the contiguous bands, leading to redundant information (Thenkabail et al., 2004) and computationally intensive processes. Therefore, a spectral reduction is required to extract relevant information and eliminate these redundancies. Principal component analysis (PCA) performs an orthogonal transformation from the initial spectral space to another space of equal dimension showing no linear correlation between latent features. These latent features (named hereafter principal components, or PCs) are then ranked, following a decreasing amount of variance explained, which is a criterion commonly used to perform component selection. Indeed, PCs explaining a low amount of variance usually contain only noise. However, the choice of selecting PCs explaining the most variance may lead to suboptimal selection for a given application, as the signal may be influenced by several factors, and those being of interest for the considered application may not lead to high variance values (contrarily, those leading to high variance values may not be of interest). It is known for instance that the influence of brightness is particularly strong on radiometric signals measured from vegetation when using high spatial resolution imagery with pixels smaller than ITCs (Fung & LeDrew, 1987; Horler & Ahern, 1986). Indeed, the angle of view, the illumination and the surface geometry are responsible for directional effects and shade. Even though brightness accounts for most of the total variance, this factor is not a relevant criterion to differentiate individuals since spectral variations due to brightness are particularly strong within individuals and may not evidence dissimilarities between ITCs. On the other hand, relevant factors for the delineation of ITCs are related to individual- or species-specific traits such as leaf chemistry (for instance, photosynthetic pigments or water content) and vegetation structure (foliage density, leaf angle distribution, tree shape, etc). These factors are known to also significantly influence spectral properties measured from individual trees (Conese et al., 1988; Morton, 1986), and this influence should be featured by some PCs. On the opposite, the selection of PCs showing irrelevant information for ITC segmentation is in the best case responsible for lower computational performances, and in the worst case a source of nuisance for the accurate delineation of ITCs.

Therefore we studied the influence of the identity of the retained PCs on the quality of tree crown segmentation. The selection or exclusion of a PC is related to the user ability to visually assess the presence of information allowing species discrimination in the PC. This information, called discriminant information, is contained in a PC whenever there are some individuals or groups of individuals clearly distinguishable from the background in the component. The following four spectral configurations were investigated:

- the initial and unprocessed hyperspectral data, showing strong correlations between bands,
- the output of the PCA transformation, without PC selection,
- a selection of PCs, visually selected to contain useful information for species discrimination. This discriminant information was visually assessed by the user, based on the presence of patterns highlighting differences between individuals, and in our case, PC #1 was not selected due to the reason explained earlier.
- the previously mentioned selection of PCs, plus PC #1.

A permissive strategy was adopted during the visual examination: components showing a few individuals were retained even if the component looked noisy overall. Even if the amount of variance was not appropriate to select components, we noticed that the interesting information was contained in the first half of all components. Figure 5 exhibits a subset of the image corresponding to the Hawaii site and its first five PCs. Discriminant information can be seen in figures 5c to 5f, where some individuals are clearly distinguishable within the components. Table 1 specifies the number of bands and the identity of the PCs used in each case for the two different sites.

Table 1: Number of bands used to perform BPT segmentation on the two study sites and identity of the component selected.

| | Hawaii | Panama |
|---|---|---|
| Hyperspectral image | 24 | 175 |
| PCA transformation | 24 | 175 |
| Visual PC selection + PC#1 | 8 | 22 |
| Visual PC selection | 7 | 21 |
| Component selected through visual inspection | 2-8 | 2,3,5,9,10,12,13,15,17-21 23,25,28,29,33,34,42,46 |

### *3.3.2. Spatial reduction*

The partition used to initialize the construction of the BPT can be composed of individual pixels (the finest partition scale) (Valero et al., 2013a), or regions obtained from a preliminary segmentation. The former is recommended when no prior information is known about the size of final regions, but the latter option is computationally more efficient as it significantly decreases the number of nodes within the BPT. In our application, the dimension of emerging ITCs ranged between tens of pixels and thousands of pixels for the largest individuals. Therefore, a preliminary segmentation of the original image was investigated for the construction of the BPT. The main constraint of this pre-segmentation was to produce regions smaller than individual trees in order to avoid grouping several ITCs in one region, as the algorithm does not include region splitting. The boundaries of the regions obtained from the pre-segmentation should also respect as much as possible the actual boundaries between ITCs in order to recompose them with a good accuracy. We investigated three different approaches to produce the initial segmentation map, and compared them with an initialization at the pixel level. Each approach used to derive the initial segmentation map was based on different initial data and different segmentation methods:

- The first approach used the LiDAR-derived CHM, as presented in figures 1b and 2b. A preliminary smoothing was first applied to the initial CHM, including the application of a 5 by 5 median filter followed by a discretization using steps of 0.5 m. This discretized CHM was then segmented using the Watershed algorithm (Beucher & Lantuejoul, 1979; Meyer & Beucher, 1990), which tends to produce strongly-over-segmented regions.

- The second approach was based on hyperspectral Watershed segmentation, as exposed in Noyel et al. (2007); Tarabalka et al. (2010). First, the gradient map of the original hyperspectral data was computed, using a Robust Color Morphological gradient (Tarabalka et al., 2010). Then, a classical Watershed algorithm was applied onto this gradient map, once again resulting in a strongly over-segmented partition.

- The third approach was based on the mean shift clustering (Comaniciu & Meer, 2002) of a RGB representation of the hyperspectral data. Bands centered at 646.0 nm (R), 560.7 nm (G) and 447.0 nm (B) were used for Hawaii, and bands centered at 638.83 nm (R), 548.77 nm (G) and 458.71 nm (B) were used for Panama. The mean shift clustering was performed with the freeware Edge Detection and Image SegmentatiON (EDISON, `http://coewww.rutgers.edu/riul/research/code/EDISON/`).

In all cases, the resulting initial segmentation maps were satisfying after visual examination, as the obtained regions were small enough to prevent several individuals to be merged in one region. Figure 6 shows the initial segmentation maps corresponding to the three methods.

### *3.4. Construction of the binary partition tree*

The construction of the BPT starts once the pre-processing step is completed, and depends on the definition of a region model and a merging criterion. There are two commonly used region models when dealing with hyperspectral images (Valero, 2011; Valero et al., 2010a). One can choose to model a hyperspectral region by its mean spectrum (also called *first order parametric region model*), which allows the use of simple merging criteria measuring the discrepancy between two spectra. However, such merging criteria proved to perform poorly when used to discriminate tree species in tropical forests (Clark et al., 2005), as they assume spectral homogeneity within each region and do not preserve their spectral distribution and variability. The *non-parametric statistical region model* is more satisfying for our application, as it accounts for spectral variability within a region. In that case, the region is modeled by its set of histograms as follows:

$$\mathcal{H}_{\mathcal{R}} = \left(\mathcal{H}_{\mathcal{R}}^{\lambda_1}, \ldots, \mathcal{H}_{\mathcal{R}}^{\lambda_M}\right) \tag{1}$$

where $\mathcal{H}_{\mathcal{R}}^{\lambda_i}$ is the empirical distribution of reflectance values for the region $\mathcal{R}$ in the band $\lambda_i$ and $M$ is the number of spectral bands in the image. Each of these $M$ histograms can then be converted in a probability density function (pdf) after normalization (so that the sum of its bins equals 1). This allows the use of metrics which measure the discrepancy between pdfs. In particular, we decided to use the diffusion distance, proposed by Ling & Okada (2006), and successfully used as a merging criterion for the BPT construction by Valero et al. (2010a). This distance, detailed in 5, is particularly robust to illumination change and allows to handle the case where a tree crown is half-lit, half shaded. The construction of the BPT is initiated by the computation of the merging criterion between each pair of neighboring regions. Each merging iteration then involves the search of the two neighboring regions that achieve the lowest pair-wise similarity among all pairs of neighboring regions in the current segmentation map. Those two regions are consequently merged. It is noteworthy that the method was programmed to favor the merging of very small regions (Calderero & Marques, 2010), in order to decrease the risk of over-segmentation and smooth the final segmentation. In practice, the average region size in the segmentation map is computed at each merging iteration, and all regions of size less than 15% of this average size are given the merging priority.

### 3.5. Pruning of the binary partition tree

After the construction of the BPT, the pruning aims at cutting off branches so the leaves of the pruned tree correspond to meaningful regions regarding the desired application. Therefore, this step is critical to achieve a proper segmentation, and our goal is to design a generic pruning strategy giving optimal ITC delineation for various forest types and image characteristics (spatial and spectral resolutions), with minimal expert parametrization. Many pruning strategies have already been investigated in the literature for classical (Salembier & Garrido, 2000) and hyperspectral BPTs (Valero et al., 2010a,b). Among the attempts made to design a generic pruning strategy, one can cite the minimization of an energy or cost function, or recursive spectral graph partitioning (Alonso-Gonzalez et al., 2013; Valero, 2011). The former associates a pruning cost to each node in the BPT and looks for partition minimizing the overall cost, subject to a given number of region in the partition, through the use of Lagrangian multipliers. This strategy requires the knowledge of the final number of regions in the image to be operated. It is inapplicable in our study as this parameter is not known *a priori*. Therefore, we propose a new pruning strategy devoted to the segmentation of tree crowns in hyperspectral images and compare its results with those obtained using the recursive spectral graph partitioning.

#### 3.5.1. Recursive spectral graph partitioning pruning strategy

The recursive pruning strategy that we use as reference and compare to our method is based on two techniques: spectral graph partitioning (Von Luxburg, 2007) and normalized cuts (Shi & Malik, 2000). This pruning strategy analyzes each branch of the BPT, seeking the best level to partition it in two sets, where the similarity among all the nodes of a given set is high, and the similarity across the two sets is low. Given that, each leaf of the BPT votes for the ancestor in the branch it wishes to be represented by. For each branch, the cut is then made under the node which has the highest ratio of votes with respect to the number of leaves hanging under it, in order not to favor nodes close to the root which have a greater number of leaves and potentially a great number of votes. The partitioning process only relies on dissimilarities among nodes of the BPT, and thus does not assume any particular knowledge about the currently processed image.

*3.5.2. The evolution of the region size pruning strategy*

The above-presented pruning strategy is based on spectral properties of graphs constructed from the BPT and depends neither on the scene depicted by the hyperspectral image nor on the application. However, it may not be optimal for such specific applications as the segmentation of tree crowns in a tropical rain forest hyperspectral images. Moreover, the solving of the graph partitioning problem can become computationally intensive for large images and potentially huge BPTs. To overcome this limitation, we propose a novel pruning strategy by adapting the aforementioned voting process to the tree crown segmentation in tropical forests. Since the initial segmentation map is over-segmented, each ITC is initially split up into several regions. Two neighboring regions belonging to the same ITC are theoretically closer spectrally than two neighboring regions belonging to two ITCs of different species. As a result, all the regions defining an ITC should have low pair-wise distances and therefore be merged in the early iterations of the merging algorithm. Those early iterations lead all regions to reach some critical size at which point their neighboring regions are spectrally dissimilar because containing one or several ITCs belonging to different species. Final iterations of the merging process usually involve regions comprising one or several individuals. As a result, the evolution of the region size from a leaf of the BPT to its root shows a clear discontinuity at the step where the region is no longer agglomerating leaves around it, but is merging instead by default with another grown up region in its neighborhood. We observed in practice that the most accurate delineation of the ITC corresponds to the region defined right before the discontinuity, as it can be observed in figure 7.

Our novel pruning strategy is derived from this observation: each leaf votes for the node prior to the first discontinuity in the branch. The introduction of a size thresholding parameter allows the detection of a discontinuity: a discontinuity is flagged when the size difference between two consecutive nodes exceeds the threshold. The pruning is decided after all leaves have voted: each non-leaf node in the BPT has its number of votes divided by its number of leaves, and each BPT branch is cut under the node whose ratio number of nodes/number of leaves is the highest in the branch. If two nodes have the same ratio in a branch, then the cut is made under the one which is the farthest apart from the root, to decrease the chance of under-segmentation. By setting the size threshold and thus controlling the discontinuity height, it is possible to influence the characteristic size of the final regions: the setting of a low threshold value tends to generate small regions since the voting process is more sensitive to leaps in the evolution of the region size. Contrarily, a high value leads to large regions in the corresponding segmentation. For Hawaii site, threshold values from 200 to 2000 with a 200 step wide have

Table 2: Basic statistics about the delineated ITCs for both test sites.

| | Hawaii | Panama |
|---|---|---|
| Number of ITCs | 160 | 100 |
| Mean size (in pixels) | 843 | 205 |
| Standard deviation | 648 | 158 |
| Minimal ITC size | 36 | 39 |
| Maximal ITC size | 3846 | 778 |

been tested, whereas for Panama site, where the average crown size is smaller, values ranging from 150 to 1500 with a 150 step wide have been tried.

### *3.6. Assessing the segmentation accuracy*

Assessing a segmentation quality is a difficult task in general, since it requires the definition of meaningful evaluation criteria, and those criteria are often to be defined with respect to a given goal and available ground truth data. Most criteria found in the literature, such as symmetric and asymmetric distances (Cardoso & Corte-Real, 2005), ask for a reference segmentation to be used. However, only some manually delineated ITCs are available in our case. Table 2 displays some basic statistics regarding those ITCs.

Once the tree has been pruned, an ITC can be described in the corresponding segmentation by one of the following four different states: *detected*, *over-segmented*, *under-segmented*, or *missed*. We propose to evaluate the segmentation accuracy by using the percentage of ITCs which were classified as correctly detected regarding the total number of ITCs tested. It is very unlikely that an automatically delineated crown exactly matches a manually delineated one. This inaccuracy between the two regions, which can be evaluated by the number of missegmented pixels, also depends on the size of the region manually delineated. Therefore, we define in the following some criteria integrating a margin of error between the manually delineated ITCs and the one obtained from the segmentation process. For a given manually delineated ITC, the first step is the retrieval of segments that represent the ITC the best in the final segmentation map. In practice, every segment that shares at least 50% of its pixels with the ITC is considered an element of the ITC. In the case where no segment has at least 50% of its pixels belonging to the ITC, then the ITC is represented by the segment having the highest percentage of pixels in it. Consequently, an ITC can be composed of one segment or several segments. In the following, $c$ denotes the set of pixels corresponding to the ITC, and $s = \{s_1, \ldots, s_{ITC}\}$ is the set of segments in the final segmentation map that were retrieved to compose the ITC. Figure 8 presents the process to determine how $c$ has to be classified regarding its corresponding set of segments $s$:

- The first test concerns the over-segmentation. The ITC crown appears to be over-segmented if several segments were found to compose it, and that case is treated aside. If $s$ contains only one segment, the *overlap* degree between $s$ and $c$ is computed. It is defined by

  $$\text{overlap} = \frac{|c \cap s|}{|c|} \tag{2}$$

  where $|c \cap s|$ denotes the number of pixels in the intersection of $c$ and $s$, and $|c|$ is the number of pixels composing the ITC. It represents how much of the ITC was captured by the segment representing it.

Consequently, if this overlap degree does not exceed at least 0.7 (the segment representing the ITC contains less than 70% of the ITC), the ITC is classified as *missed*.

- If the ITC is not missed, the *ratio* degree between $s$ and $c$ is computed,

$$\text{ratio} = \frac{|s|}{|c|} \tag{3}$$

  If this ratio is greater than 1.5 (the segment is at least 50% bigger than the ITC it represents), then the ITC is classified as *under-segmented*.

- If the ITC is neither missed nor under-segmented, then it is classified as *detected*.

- In the case were the ITC was found to be over-segmented, an additional test examines how severe is the over-segmentation. If there is one segment $s^\star \in \{s_1, \ldots, s_{ITC}\}$ such that $s^\star$ accounts for at least 85% of the area covered by $s$, and the *overlap* and *ratio* degrees of $s^\star$ alone are such that they makes the ITC being detected, then the over-segmentation is discarded and the ITC is classified as *detected*. Otherwise, the ITC remains *over-segmented*.

All the previous cases can be observed in figure 9. Threshold values for overal and ratio degrees and to discard over-segmentation were set empirically. The influence on the final segmentation quality of each input parameter (the initial segmentation map and the PCA configuration) and of the pruning strategy can be assessed by the percentage of correctly delineated ITCs.

## 4. Results and discussion

### *4.1. Results*

Table 3: Percentage of correctly segmented ITCs for Hawaii test site, according to the chosen setting. A setting is defined by a spectral reduction configuration (No PCA, all PCs, Selection of PCs), an initial segmentation (pixel scale, mean shift clustering, hyperspectral watershed, watershed on LiDAR) and a pruning strategy (graph cut, region size discontinuity). For the region size discontinuity pruning strategy, several threshold values were investigated: is reported the maximum percentage along with the corresponding threshold value (in parentheses).

| **HAWAII** | | No PCA | All PCs | Selection of PCs | |
|---|---|---|---|---|---|
| | | | | with 1$^{st}$ PC | without 1$^{st}$ PC |
| *Graph cut* | Pixel Scale | **15.0** | **24.4** | **28.1** | **33.8** |
| | Mean Shift | **32.5** | **38.8** | **40.0** | **42.5** |
| *Region size discontinuity* | Hyperspectral | **6.9** (600) | **30.6** (1400) | **29.4** (1200) | **40.0** (1600) |
| | LiDAR | **36.9** (600) | **47.5** (600) | **47.5** (600) | **48.8** (600) |
| | Mean Shift | **28.1** (1000) | **47.5** (1000) | **45.6** (1600) | **54.4** (1200) |

Table 3 and 4 display the percentages of ITCs correctly delineated for the Hawaii and Panama test sites, respectively, with respect to varying input parameters and pruning strategies. The two investigated pruning strategies are denoted *graph cut* for the recursive spectral graph partitioning strategy, and *region size discontinuity* for the proposed evolution of the region size strategy. The initial segmentation maps are denoted as follows:

Table 4: Percentage of correctly segmented ITCs for Panama test site, according to the chosen setting. A setting is defined by a spectral reduction configuration (No PCA, all PCs, Selection of PCs), an initial segmentation (pixel scale, mean shift clustering, hyperspectral watershed, watershed on LiDAR) and a pruning strategy (graph cut, region size discontinuity). For the region size discontinuity pruning strategy, several threshold values were investigated: is reported the maximum percentage along with the corresponding threshold value (in parentheses).

| **PANAMA** | | No PCA | All PCs | Selection of PCs | |
|---|---|---|---|---|---|
| | | | | with $1^{st}$ PC | without $1^{st}$ PC |
| *Graph cut* | Pixel Scale | **44.0** | **54.0** | **51.0** | **63.0** |
| | Mean Shift | **54.0** | **54.0** | **59.0** | **61.0** |
| *Region size discontinuity* | Hyperspectral | **26.0** (150) | **33.0** (150) | **42.0** (150) | **43.0** (450) |
| | LiDAR | **39.0** (150) | **55.0** (150) | **51.0** (150) | **49.0** (150) |
| | Mean Shift | **45.0** (150) | **63.0** (150) | **66.0** (150) | **68.0** (150) |

*pixel scale* when the initialization is done at pixel level, *hyperspectral* for the hyperspectral Watershed segmentation, *LiDAR* for the classical Watershed algorithm applied on LiDAR data, and *mean shift* for the mean shift clustering. The input images on which the BPT is built are denoted as *No PCA*, *All PCs*, *Selection of PCs with* $1^{st}$ *PC* and *Selection of PCs without* $1^{st}$ *PC* for the raw hyperspectral data, the PCA transformation with all PCs retained, the PCA transformation with only a selection of PCs, with and without PC#1 retained, respectively. Several threshold values were tested for the proposed pruning strategy, ranging from 200 to 2000 with gaps of 200 for

Hawaii, and from 150 to 1500 with gaps of 150 for Panama. The maximum percentage along with its corresponding threshold value is reported. The highest percentage among all configurations is highlighted in red. For both sites, this configuration involves a spectral reduction performed by PCA with a selection of PCs excluding PC #1, an initial segmentation resulting from the mean shift clustering and the proposed region size discontinuity strategy for the pruning of the BPT. Rationales of this conclusion are discussed in the next subsection 4.2. Figure 10 displays some segmentation results obtained for both sites.

## *4.2. Discussion*

### *4.2.1. About the PCA configuration*

For both test sites, all initial partitions and both pruning strategies, results show significant improvements when a PCA transformation is performed with respect to the case where the BPT is directly built on the raw hyperspectral data. We conclude that the discriminant information extracted by the PCA eases the species discrimination and improves the region model for the BPT. Regarding the most efficient PCA configuration, there is no clear conclusion about the best strategy between using all PCs and using only the selection of PCs including PC #1. There is in fact very little variations in terms of amount of discriminant information between those two configurations, since all bands exluded in the latter configuration contained only and no useful information for species discrimination. On the other hand, discarding the first PC improves the results. As explained in section 3.3.1, the first PC contains brightness variations measured in the NIR domain, which is a highly non-discriminative feature. Including this first PC can be prejudicial for instance when a tree crown is half-lit, half-shaded. In that case, the distance between the two halves increases during the construction of the BPT as their histograms corresponding to the first PC show significant differences. The two halves may even not be merged together, preventing the tree crown to be correctly segmented during the pruning step. As a result, the selection of PCs without PC #1 included is the configuration which gives the highest percentage of detected ITCs among all studied spectral reduction configurations.

#### *4.2.2. About the initial segmentation map*

Among the three pre-segmentation methods investigated, the hyperspectral Watershed systematically gives the lowest percentage of correctly segmented ITCs. This is counter-intuitive since the hyperspectral Watershed produces smaller regions than the two other segmentation methods (see figure 6), hence should decrease the risk that its regions already group several ITCs together. However, when precisely analyzing those initial regions, one can see that they all have the same square-like shape. On the opposite, initial regions derived from LiDAR Watershed and mean shift clustering are more heterogeneous in shape and size and have more pertinent boundaries (some ITCs are directly recognizable in the initial segmentation map). This is plausibly due to the fact that the hyperspectral Watershed involves the computation of a multidimensional gradient on the raw hyperspectral data. The noise present in this data, along with a low ground sampling resolution, leads to an imprecise gradient map. The following Watershed on this gradient map therefore produces regions not only following the real edges of the image (high value areas in the gradient map), but also fake edges introduced by the noise, thus initial regions lacking consistency. In contrast, mean shift clustering and LiDAR Watershed are based on three bands of the raw hyperspectral data and on a smoothed version of the CHM, respectively. They are consequently less sensitive to noise and generate more accurate regions. This emphasizes the necessity for the initial segmentation map to have regions relevant enough to recompose accurately the real boundaries between ITCs. Additionally, the LiDAR Watershed method tends to produce larger initial regions than the mean shift clustering method (if several neighboring trees have the same height, they will likely be grouped in the same region for instance), increasing the risk of final under-segmentation. For the Panama site, ITCs have a rather small size (205 pixels in average for the 100 delineated ITCs) and are consequently more sensible to under-segmentation, explaining why the LiDAR Watershed is slightly outperformed by the mean shift clustering. This is less true for Hawaii test site where ITCs have a larger size and where both segmentation methods produce comparable results.

Conversely, there are many more possible merging combinations when the BPT is initialized on the pixel level, therefore more chances to miss or over-segment a region. It is in fact easier to reconstruct a real region when its borders are already partially known, as it is the case if the initial segmentation was correctly performed. This is particularly true when the expected regions have a rather large size, explaining why mean shift clustering leads to better results than the pixel scale for Hawaii site. However, both mean shift clustering and pixel scale initialization give similar results for Panama, due to smaller regions. Nevertheless, the number of nodes in the BPT is proportional to the number of regions in the initial partition. That is why the use of mean shift clustering as an initial pre-segmentation should still be preferred as it drastically decreases the number of regions in the initial partition (thus reducing the computational load) without degrading the results.

### *4.2.3. About the pruning strategy*

It is more challenging to compare the performances of the two investigated pruning strategies. As said in section 3.5, the pruning strategy strongly depends on the application. The recursive spectral graph partitioning strategy tries to be as generic as possible, only exploiting dissimilarities along each branch of the BPT, for every type of images. Our pruning strategy, presented in section 3.5.2 relies on a property holding when a BPT is built on an image which contains regions with a limited size range. This is indeed the case for forested areas since real regions correspond to tree crowns, which have an upper and lower bound in size for physical reasons, ensuring a clear discontinuity in the evolution of the region size along the corresponding BPT branch. When analyzing detection percentages, it can be seen that our proposed pruning strategy leads to slightly better results than the recursive spectral graph cut pruning strategy, confirming that it is more appropriate for the segmentation of tree crowns.

### *4.2.4. About the threshold value for the proposed pruning strategy*

The tuning of the threshold value for the proposed pruning strategy is also an important point. As said in section 3.5.2, the threshold value impacts the average region size in the final segmentation map. Indeed, a high threshold value is permissive in terms of discontinuity in the evolution of the region size along a branch since larger discontinuities are allowed. Consequently, leaves vote for nodes closer to the root, hence large final regions and a potential under-segmentation of the image. On the other hand, a low threshold value is sensitive in terms of discontinuity, and favors small regions in the final segmentation while increasing the chances of over-segmentation. Naturally, the percentage of over-segmented (under-segmented) ITCs is a decreasing (an increasing) function of the threshold value, as it can be observed in figure 11. On the other hand, the percentage of missed ITC remains relatively constant (an ITC is declared missed when there is no region matching it). It is then clear that a threshold value can be considered optimal when it achieves a trade-off between over-segmentation and under-segmentation phenomena. There is no explicit rule to find the best value achieving such compromise, but one can remark that it should be close to the average size of expected regions. As a matter of fact, figure 11a shows that threshold values achieving the best trade-off between over- and under-segmentation for Hawaii, PC selection without PC #1 and mean shift clustering are 1000 and 1200 whereas table 2 exhibits a mean ITC size of 843 pixels. For Panama, figure 11b gives optima threshold values of 150 and 300 while the average ITC size is 205. The difference regarding the average ITC size between the two sites can be explained by i) the difference in spatial resolution between the two images (0.56 m for Hawaii and 2 m for Panama), and ii) the structural differences of individual trees between these two sites, explained by physical, environmental and anthropic factors. Therefore, one can roughly estimate a threshold value based on the average size of the expected regions (regarding the characteristics of the image to segment), and then adapt this value depending on the result, if needed. A means to locally and automatically adjust the threshold value would overcome the supervised nature of the method as well as ensuring robustness regarding a highly variable ITC size.

#### 4.2.5. About the general performances of the proposed method

Tropical rainforests are among the richest and most complex ecosystems in the world. Given the density of the canopy in terms of individuals and species, as well of the complexity of its structure, achieving a perfect delineation of each tree crown is highly unrealistic. However, even partial information allowing a better delimitation, identification and enumeration of certain species of interest (such that dominant, rare or invasive species that are key indicators for environmental processes) can help ecologists to better understand these complex ecosystems. Despite several studies about tree crown classification of tropical rainforests (see for example Feret & Asner (2013) or Clark et al. (2005)), there is, to best of our knowledge, no reference study for the segmentation of tropical rainforests. Bunting & Lucas (2006) developed a segmentation method for hyperspectral images, and applied it on Compact Airborne Spectrographic Imager (CASI) data acquired over mixed Australian forests. They reported over 70% of success for the segmentation of trees or clusters of trees belonging to the same species, for relatively sparse vegetation covers. However, they noted a significant drop in this segmentation accuracy for dense and complex canopies. Results obtained by our proposed method (up to 54.4 % for Hawaii and 68% for Panama in the best cases) for the delineation of tree crowns with various characteristics (such as size, shape or species) are therefore very promising. Moreover, segmentation results are visually consistent, as can be seen in figure 10. This motivates us to pursue additional measures to improve the proposed method, in order to better identify and segment tree crowns in tropical rainforests.

## 5. Conclusion

The accurate and automatic delineation of tree crowns in tropical rainforests allows application of various object-oriented methods, for example the estimation of leaf chemistry, and tree species identification which proved to perform better than pixel-oriented counterparts. However this task is extremely challenging in these complex ecosystems. Here, we presented a method for the segmentation of hyperspectral images of tropical rainforests, based on binary partition trees. The evaluation of our method was conducted on two test sites presenting different image properties (ground sampling distance and number of spectral bands) and forestry characteristics. The contributions of the present study are the following:

- The adaptation of the generic BPT algorithm to a specific application, being the segmentation of tree crowns in hyperspectral images of tropical rainforests. This was done through the selection of pertinent region model and merging criterion.

- The introduction of a pre-processing step including spectral and spatial dimensionality reduction. The former, achieved using a PCA transformation, demonstrated how the PCA extracts and highlights discriminant information when applied on images acquired over forested covers. It also illustrated the low discriminant capacity of the first PC by comparing several PC combinations as the input image for the BPT construction. The latter showed the interest of initializing the BPT on an initial over-segmentation of the image with respect to the pixel level. We showed how this pre-segmentation has to meet strict requirements in terms of size and borders of the generated regions. The results of three different segmentation algorithms were compared. Mean shift clustering proved to be the most efficient method among the three investigated.

- The introduction of a new BPT pruning strategy, based on a voting process where each leaf of the BPT elects its favorite ancestor. The vote depends on the evolution of the region size along a branch, as we remarked a clear discontinuity in terms of region size for the node whose corresponding region represents a tree crown the best. Not only this pruning strategy is adapted for the segmentation of forested areas, but also for images featuring a patchwork of homogeneous regions. We compared this novel pruning strategy with an already existing one, based on spectral graph partitioning. Results showed that the proposed pruning strategy was more adapted to this precise task.

- The introduction of a method assessing the segmentation quality, based on the knowledge of some reference regions only. Indeed, due to the high complexity of the canopy, it is unrealistic to generate a reference segmentation manually. To overcome this issue, ITCs were manually delineated and accounted for ground-truth. A particular care was taken to select ITCs of various sizes and shapes, and representing the species diversity. We proposed to classify these ITCs into four categories depending on their segmentation state, namely correctly detected, over-segmented, under-segmented and missed. The segmentation quality was then defined as the percentage of ITCs correctly segmented.

We are now working on using LiDAR data in a more optimal way. As for now, LiDAR was only used to provide an initial segmentation map, the BPT being built on the raw or transformed hyperspectral data, thus relying only on spectral properties of the scene. However, by incorporating the LiDAR during the BPT construction, physical properties such as the height or diameter of the crown could be taken into consideration. In particular, the use of LiDAR could overcome the case where several trees of the same species are aggregated together and are likely to appear as only one region if using only spectral properties. The automated selection of PCs containing discriminant information as well as the automated tuning of the threshold value for the BPT pruning will also be investigated in order to make the proposed method fully unsupervised.

**Acknowledgements**

The presented work has been done in a joint collaboration between the GIPSA-Lab laboratory at Grenoble (France) and the Department of Global Ecology, Carnegie Institution for Science at Stanford (USA). Within the GIPSA-Lab, this work has been supported by project XIMRI, ANR 2010 INTB 0208 01. The Carnegie Airborne Observatory is made possible by the Gordon and Betty Moore Foundation; the Grantham Foundation for the Protection of the Environment; the John D. and Catherine T. MacArthur Foundation; the Avatar Alliance Foundation; the W. M. Keck Foundation; the Margaret A. Cargill Foundation; Mary Anne Nyburg Baker and G. Leonard Baker Jr.; and William R. Hearst III.

## Appendix A. About the diffusion distance as a BPT merging criterion

Detailed below is the expression of the diffusion distance when used as a merging criterion for the construction of a BPT with a non-parametric statistical region model. This distance was proposed by Ling & Okada (2006) as a measure of discrepancy between histograms. The underlying idea is to view the difference between two histograms as a temperature field. The distance between the two histograms is based on the time needed for the temperature distribution to reach stability via a heat diffusion process, or equivalently, on the state of the temperature field after a given time. Opposed to *bin-to-bin* distances which assume that histograms are already aligned and compare a bin in one histogram only to the corresponding bin in the other histogram, the diffusion distance is a *cross-bin* distance and is usable even when histograms are not aligned. More specifically, for two histograms $\mathcal{H}_1$ and $\mathcal{H}_2$ whose $P$ bins are denoted by

$$a_p \quad \forall p \in [1 : P], \tag{A.1}$$

the diffusion distance first defines the difference histogram:

$$d_0(a_p) = \mathcal{H}_1(a_p) - \mathcal{H}_2(a_p), \tag{A.2}$$

and then simulate the temperature diffusion process by convolving the current temperature field with a Gaussian kernel

$$d_m(a_p) = [d_{m-1}(a_p) * g_\sigma(a_p)]\downarrow_2 \quad \forall m \in [1 : L] \tag{A.3}$$

where $g_\sigma(x)$ stands for a Gaussian kernel with variance $\sigma$, $L$ is the number of layers in the convolution process (the time after which the diffusion is stopped), and $\downarrow_2$ denotes a downsampling by factor 2. The distance between the two histograms is then obtained by summing up the $\mathcal{L}_1$ norm of each layer:

$$\mathcal{O}\left(\mathcal{H}_1, \mathcal{H}_2\right) = \sum_{m=0}^{L} \|d_m\|_1 \tag{A.4}$$

with

$$\|d_m\|_1 = \sum_{p=1}^{P} |d_m(a_p)| \tag{A.5}$$

The diffusion distance was successfully adapted to the construction of BPTs by Valero et al. (2010a). Being $\mathcal{R}_i$ and $\mathcal{R}_j$ two neighboring regions during the BPT construction, and $\mathcal{H}_{\mathcal{R}_i} = \left(\mathcal{H}_{\mathcal{R}_i}^{\lambda_1}, \ldots, \mathcal{H}_{\mathcal{R}_i}^{\lambda_M}\right)$ and $\mathcal{H}_{\mathcal{R}_j} = \left(\mathcal{H}_{\mathcal{R}_j}^{\lambda_1}, \ldots, \mathcal{H}_{\mathcal{R}_j}^{\lambda_M}\right)$ their respective region models, the diffusion distance measures for each spectral band $\lambda_k$ the similarity between the pair of histograms $\mathcal{H}_{\mathcal{R}_i}^{\lambda_k}$ and $\mathcal{H}_{\mathcal{R}_j}^{\lambda_k}$, $\mathcal{O}\left(\mathcal{H}_{\mathcal{R}_i}^{\lambda_k}, \mathcal{H}_{\mathcal{R}_j}^{\lambda_k}\right)$. The merging criterion between the two regions $\mathcal{R}_i$ and $\mathcal{R}_j$ immediately follows on by adding up the contribution of the $M$ spectral bands:

$$\mathcal{O}\left(\mathcal{R}_i, \mathcal{R}_j\right) = \sum_{k=1}^{M} \mathcal{O}\left(\mathcal{H}_{\mathcal{R}_i}^{\lambda_k}, \mathcal{H}_{\mathcal{R}_j}^{\lambda_k}\right) \tag{A.6}$$

## References

Alonso-Gonzalez, A., Valero, S., Chanussot, J., Lopez-Martinez, C., & Salembier, P. (2013). Processing multidimensional sar and hyperspectral images with binary partition tree. *Proceedings of the IEEE*, *101*, 723–747.

Andersen, H.-E. (2003). *Estimation of critical forest structure metrics through the spatial analysis of airborne laser scanner data*. Ph.D. thesis University of Washington.

Asner, G. P., Jones, M. O., Martin, R. E., Knapp, D. E., & Hughes, R. F. (2008). Remote sensing of native and invasive species in hawaiian forests. *Remote Sensing of Environment*, *112*, 1912–1926.

Asner, G. P., Knapp, D. E., Boardman, J., Green, R. O., Kennedy-Bowdoin, T., Eastwood, M., Martin, R. E., Anderson, C., & Field, C. B. (2012). Carnegie airborne observatory-2: Increasing science data dimensionality via high-fidelity multi-sensor fusion. *Remote Sensing of Environment*, *124*, 454–465.

Asner, G. P., Knapp, D. E., Broadbent, E. N., Oliveira, P. J., Keller, M., & Silva, J. N. (2005). Selective logging in the brazilian amazon. *Science*, *310*, 480–482.

Asner, G. P., Knapp, D. E., Kennedy-Bowdoin, T., Jones, M. O., Martin, R. E., Boardman, J., & Field, C. B. (2007). Carnegie airborne observatory: in-flight fusion of hyperspectral imaging and waveform light detection and ranging for three-dimensional studies of ecosystems. *Journal of Applied Remote Sensing*, *1*, 013536–013536.

Asner, G. P., Rudel, T. K., Aide, T. M., Defries, R., & Emerson, R. (2009). A contemporary assessment of change in humid tropical forests. *Conservation Biology*, *23*, 1386–1395.

Beucher, S., & Lantuejoul, C. (1979). Use of watersheds in contour detection. In *International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation, Rennes, France.*.

Bunting, P., & Lucas, R. (2006). The delineation of tree crowns in australian mixed species forests using hyperspectral compact airborne spectrographic imager (casi) data. *Remote Sensing of Environment*, *101*, 230–248.

Calderero, F., & Marques, F. (2010). Region merging techniques using information theory statistical measures. *Image Processing, IEEE Transactions on*, *19*, 1567–1586.

Cardoso, J. S., & Corte-Real, L. (2005). Toward a generic evaluation of image segmentation. *Image Processing, IEEE Transactions on*, *14*, 1773–1782.

Clark, M. L., Roberts, D. A., & Clark, D. B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote sensing of environment*, *96*, 375–398.

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *24*, 603–619.

Conese, C., Maracchi, G., Miglietta, F., Maselli, F., & Sacco, V. (1988). Forest classification by principal component analyses of tm data. *International Journal of Remote Sensing*, *9*, 1597–1612.

Culvenor, D. S. (2002). Tida: an algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery. *Computers & Geosciences*, *28*, 33–44.

Drumetz, L., Veganzones, M. A., Marrero, R., Tochon, G., Dalla Mura, M., Plaza, A., Chanussot, J. et al. (2014). Binary partition tree-based local spectral unmixing. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 7th International IEEE Workshop on*.

Erikson, M. (2004). Species classification of individually segmented tree crowns in high-resolution aerial images using radiometric and morphologic image measures. *Remote Sensing of Environment*, *91*, 469–477.

Erikson, M., & Olofsson, K. (2005). Comparison of three individual tree crown detection methods. *Machine Vision and Applications*, *16*, 258–265.

Feret, J., & Asner, G. (2013). Tree species discrimination in tropical forests using airborne imaging spectroscopy. *Geoscience and Remote Sensing, IEEE Transactions on*, *51*, 73–84.

Fung, T., & LeDrew, E. (1987). Application of principal components analysis to change detection. *Photogrammetric engineering and remote sensing*, *53*, 1649–1658.

Garrido, L. (2002). *Hierarchical region based processing of images and video sequences: application to filtering, segmentation and information retrieval*. Ph.D. thesis Universitat Politnica de Catalunya-Department of Signal theory and Communications, Barcelona, Spain (April 2002).

Gougeon, F. A. (1995). A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian journal of remote sensing*, *21*, 274–284.

Horler, D., & Ahern, F. (1986). Forestry information content of thematic mapper data. *International Journal of Remote Sensing*, *7*, 405–428.

Jung, J., Pasolli, E., Prasad, S., Tilton, J., & Crawford, M. (2014). A framework for land cover classification using discrete return lidar data: Adopting pseudo-waveform and hierarchical segmentation. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *7*, 491–502.

Ke, Y., & Quackenbush, L. J. (2011). A comparison of three methods for automatic tree crown detection and delineation from high spatial resolution imagery. *International Journal of Remote Sensing*, *32*, 3625–3647.

Leckie, D. G., Gougeon, F. A., Tinis, S., Nelson, T., Burnett, C. N., & Paradine, D. (2005). Automated tree recognition in old growth conifer stands with high resolution digital imagery. *Remote Sensing of Environment*, *94*, 311–326.

Leckie, D. G., Gougeon, F. A., Walsworth, N., & Paradine, D. (2003). Stand delineation and composition estimation using semi-automated individual tree crown analysis. *Remote Sensing of Environment*, *85*, 355–369.

Ling, H., & Okada, K. (2006). Diffusion distance for histogram comparison. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* (pp. 246–253). IEEE volume 1.

Meyer, F., & Beucher, S. (1990). Morphological segmentation. *Journal of visual communication and image representation*, *1*, 21–46.

Morton, A. (1986). Moorland plant community recognition using landsat mss data. *Remote sensing of environment*, *20*, 291–298.

Noyel, G., Angulo, J., & Jeulin, D. (2007). Morphological segmentation of hyperspectral images. *Image Anal. Stereol*, *26*, 101–109.

Olofsson, K. (2002). Detection of single trees in aerial images using template matching. In *ForestSAT Symposium, Heriot Watt University, Edinburgh*.

Papes, M., Tupayachi, R., Martinez, P., Peterson, A., Asner, G., & Powell, G. (2013). Seasonal variation in spectral signatures of five genera of rainforest trees. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *6*, 339–350.

Perrin, G., Descombes, X., & Zerubia, J. (2005). *Point processes in forestry : an application to tree crown detection*. Technical Report 5544 INRIA.

Phinn, S., Ticehurst, C., Held, A., Scarth, P., Nightingale, J., & Johansen, K. (2008). New tools for monitoring world heritage values. *Living in a Dynamic Tropical Forest Landscape*, (pp. 591–609).

Pollock, R. (1996). *The automatic recognition of individual trees in aerial images of forests based on a synthetic tree crown image model*. Ph.D. thesis The University of British Columbia.

Pollock, R. (1998). Individual tree recognition based on a synthetic tree crown image model. In *Proc. of the International Forum on Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry* (pp. 25–34). Victoria, British Columbia, Canada.

Pouliot, D., King, D., Bell, F., & Pitt, D. (2002). Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sensing of Environment*, *82*, 322–334.

Pouteau, R., & Stoll, B. (2012). Svm selective fusion (self) for multi-source classification of structurally complex tropical rainforest. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *5*, 1203–1212.

Rasi, R., Beuchle, R., Bodart, C., Vollmar, M., Seliger, R., & Achard, F. (2013). Automatic updating of an object-based tropical forest cover classification and change assessment. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *6*, 66–73.

Reiche, J., Souza Jr, C. M., Hoekman, D. H., Verbesselt, J., Persaud, H., & Herold, M. (2013). Feature level fusion of multi-temporal alos palsar and landsat data for mapping and monitoring of tropical deforestation and forest degradation. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *6*, 2159–2173.

Salembier, P., & Garrido, L. (2000). Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *Image Processing, IEEE Transactions on*, *9*, 561–576.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *22*, 888–905.

Somers, B., & Asner, G. P. (2013). Invasive species mapping in hawaiian rainforests using multi-temporal hyperion spaceborne imaging spectroscopy.

*Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *6*, 351–359.

Tarabalka, Y., Chanussot, J., & Benediktsson, J. A. (2010). Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition*, *43*, 2367–2379.

Tarabalka, Y., Tilton, J. C., Benediktsson, J. A., & Chanussot, J. (2012). A marker-based approach for the automated selection of a single segmentation from a hierarchical set of image segmentations. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *5*, 262–272.

Thenkabail, P. S., Enclona, E. A., Ashton, M. S., & Van Der Meer, B. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote sensing of environment*, *91*, 354–376.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F., De Siqueira, M. F., Grainger, A., Hannah, L. et al. (2004). Extinction risk from climate change. *Nature*, *427*, 145–148.

Valero, S. (2011). *Hyperspectral image processing and representation using Binary Partition Trees*. Ph.D. thesis Gipsa-Lab, Department of Images and Signals, Grenoble Institute of Technology, Grenoble, France.

Valero, S., Salembier, P., & Chanussot, J. (2010a). Comparison of merging orders and pruning strategies for binary partition tree in hyperspectral data. In *Image Processing (ICIP), 2010 17th IEEE International Conference on* (pp. 2565–2568). IEEE.

Valero, S., Salembier, P., & Chanussot, J. (2010b). New hyperspectral data representation using binary partition tree. In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International* (pp. 80–83). IEEE.

Valero, S., Salembier, P., & Chanussot, J. (2011a). Hyperspectral image segmentation using binary partition trees. In *Image Processing (ICIP), 2011 18th IEEE International Conference on* (pp. 1273–1276). IEEE.

Valero, S., Salembier, P., & Chanussot, J. (2013a). Hyperspectral image representation and processing with binary partition trees. *IEEE Transactions on Image Processing*, *22*, 1430–1443.

Valero, S., Salembier, P., & Chanussot, J. (2013b). Object recognition in urban hyperspectral images using binary partition tree representation. In *IGARSS* (pp. 4098–4101).

Valero, S., Salembier, P., Chanussot, J., & Cuadras, C. M. (2011b). Improved binary partition tree construction for hyperspectral images: application to object detection. In *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International* (pp. 2515–2518). IEEE.

Varekamp, C., & Hoekman, D. (2001). Segmentation of high-resolution insar data of a tropical forest using fourier parameterized deformable models. *International Journal of Remote Sensing*, *22*, 2339–2350.

Veganzones, M., Tochon, G., Dalla-Mura, M., Plaza, A., & Chanussot, J. (2014). Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation. *Image Processing, IEEE Transactions on*, .

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, *17*, 395–416.

Wang, L., Gong, P., & Biging, G. S. (2004). Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery. *Photogrammetric Engineering & Remote Sensing*, *70*, 351–357.

Warner, T. A., Lee, J. Y., & McGraw, J. B. (1998). Delineation and identification of individual trees in the eastern deciduous forest. In *Proc. International Forum: Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry, Canadian Forest Service, Pacific Forestry Centre Victoria, British Columbia* (p. 8191).

Whiteside, T., & Ahmadb, W. (2008). Estimating canopy cover from eucalypt dominant tropical savanna using the extraction of tree crowns from very high resolution imagery. *GEOBIA 2008–GEOgraphic Object Based Image Analysis for the 21st century*, .

Whitmore, T. C. et al. (1990). *An introduction to tropical rain forests.*. Clarendon Press.

Wulder, M., Niemann, K. O., & Goodenough, D. G. (2000). Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sensing of environment*, *73*, 103–114.

Zhou, J., Proisy, C., Descombes, X., Hedhli, I., Barbier, N., Zerubia, J., Gastellu-Etchegorry, J.-P., & Couteron, P. (2010). Tree crown detection in high resolution optical and lidar images of tropical forest. In *Remote Sensing* (pp. 78240Q–78240Q). International Society for Optics and Photonics.

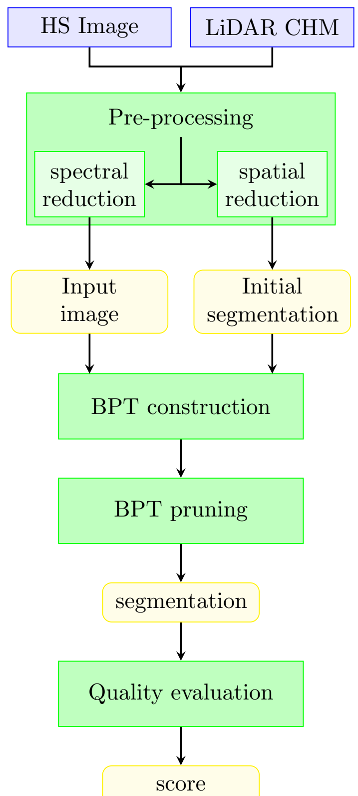



Figure 4: Flowchart of the proposed method. Blue, green and yellow rectangles correspond to input data, global operations that are further described in section 3, and outputs of those global operations, respectively.

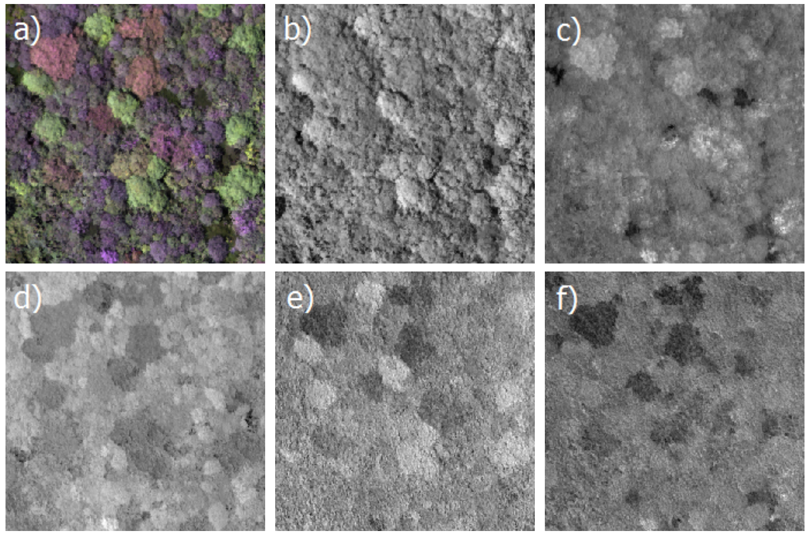



Figure 5: (a) sub-image of Hawaii site (same bands and color stretching used as in Fig. 1 for RGB representation). (b)-(f) corresponding first five principal components.

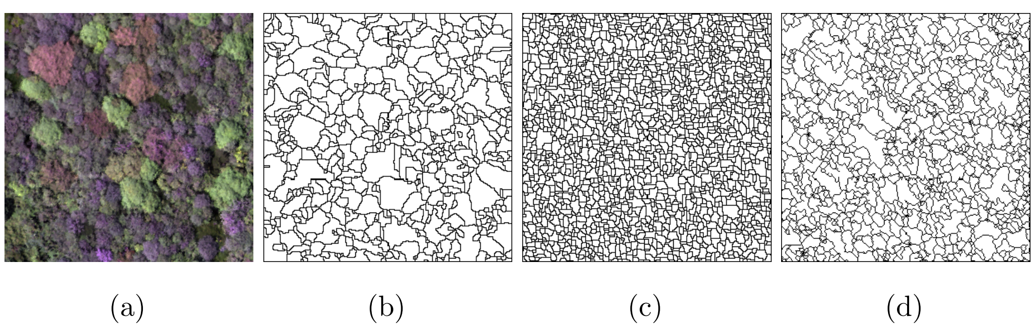
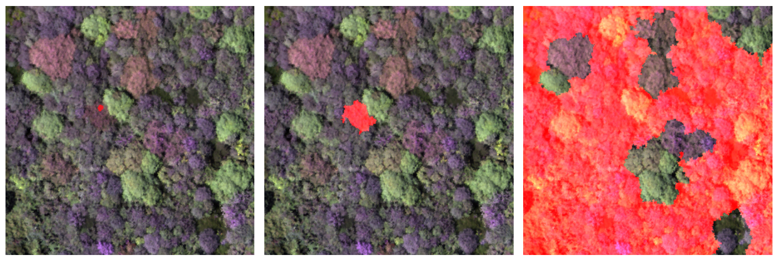
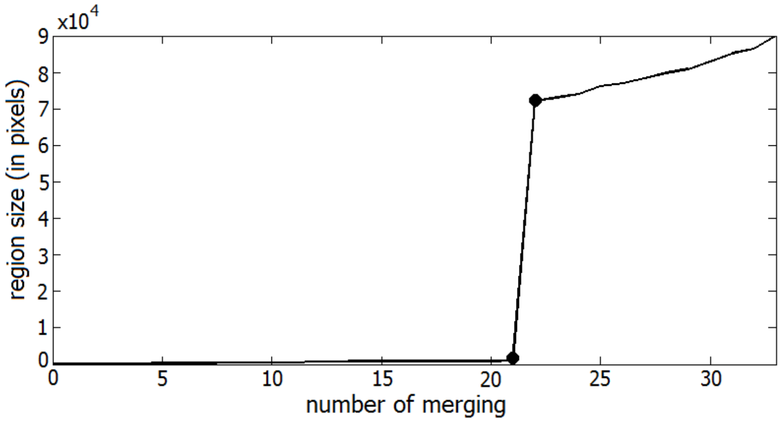



Figure 6: (a) sub-image of Hawaii site and corresponding initial segmentation using (b) Watershed algorithm applied to LiDAR CHM, (c) hyperspectral Watershed, and (d) mean shift clustering.



(a) (b) (c)




(d)

Figure 7: Evolution of a region (underlined in red) along a branch of the BPT : (a) initial region/leaf, (b) region after 21 mergings, (c) region after 22 mergings, and (d) plot of the corresponding evolution of the region size along the branch. The first and second dots correspond to the regions after 21 and 22 mergings, respectively.

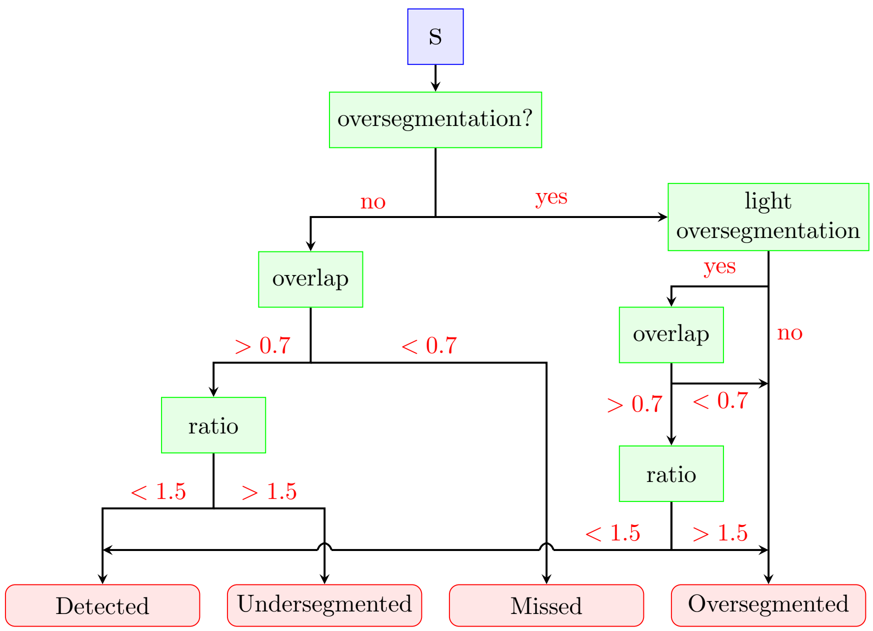



Figure 8: Flowchart summarizing the quality assessment method

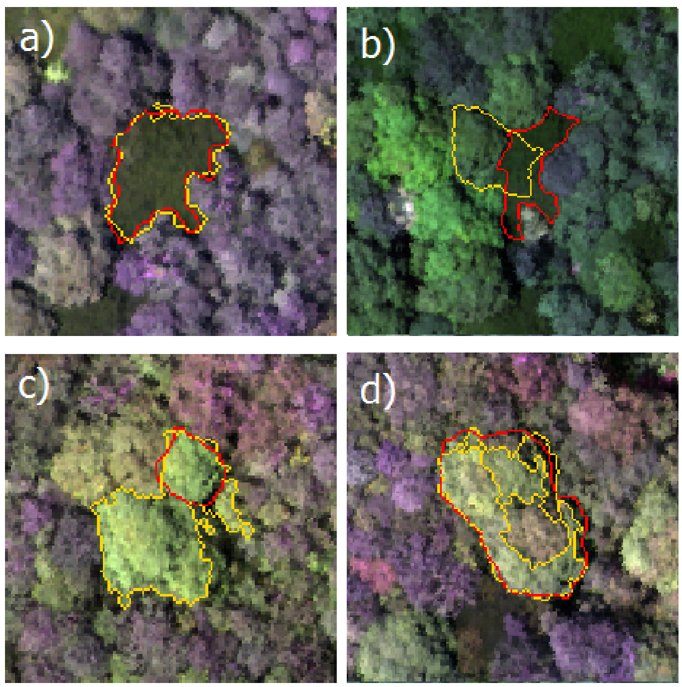
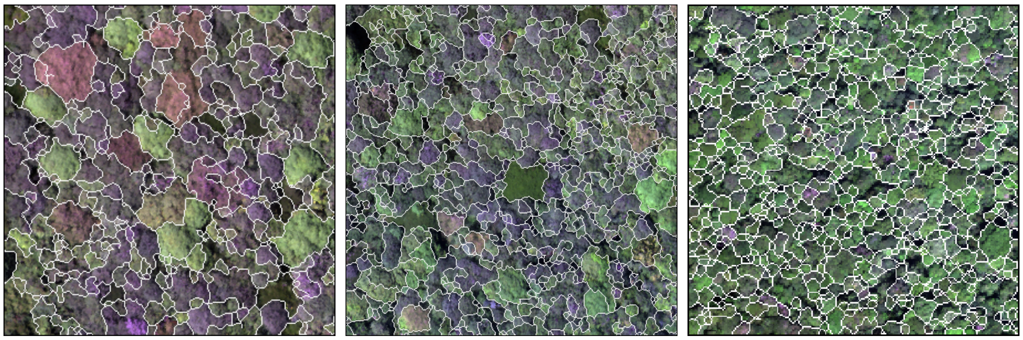
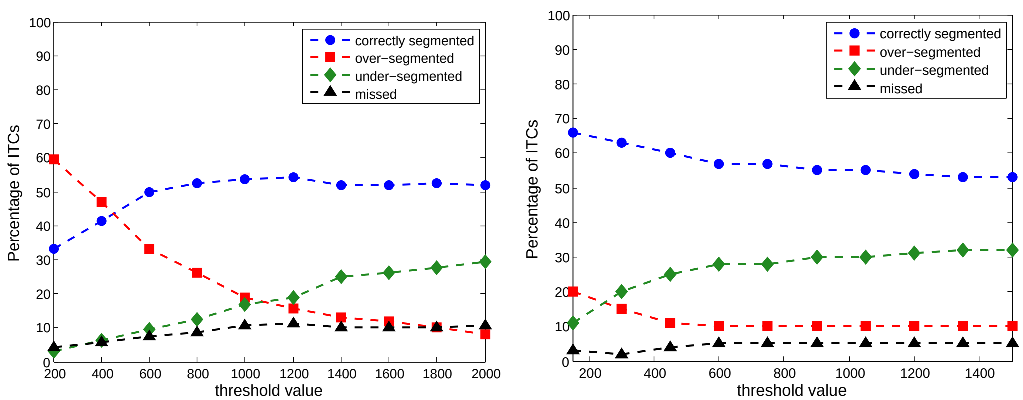



Figure 9: Manually delineated ITC (in red borders) and segmentation result (in yellow borders) for the case: (a) correctly delineated, (b) missed, (c) under-segmented, (d) over-segmented.



(a) (b) (c)

Figure 10: Visual results obtained when using mean shift clustering, PC selection without PC #1 and size threshold of 1200 for Hawaii (a,b) and 150 for Panama (c).




(a) (b)

Figure 11: Percentages of ITCs correctly segmented, over-segmented, under-segmented and missed with respect to the threshold value. Results are for (a) Hawaii site and (b) Panama site, PC selection without PC #1 and mean shift clustering.