

3D SEGMENTATION OF FOREST STRUCTURE USING A MEAN-SHIFT BASED ALGORITHM

António Ferraz^{1,2,3}, Frédéric Bretar¹, Stéphane Jacquemoud², Gil Gonçalves³, Luisa Pereira⁴

1–IGN, Institut Geographique National, MATIS Laboratory, Saint Mandé, FRANCE

2–IPGP, Institut de Physique du Globe de Paris, Space and Planetary Geophysics, Paris, FRANCE

3–INESCC, Instituto de Engenharia de Sistemas e Computadores de Coimbra, Coimbra, PORTUGAL

4–UA, Universidade de Aveiro, Escola Superior de Tecnologia e Gestão de Águeda, Águeda, PORTUGAL

E-mail: antonio.ferraz@ign.fr

ABSTRACT

Consistent and accurate information on 3D forest canopy structure is required by many applications like forest inventory, management, logging, fuel mapping, habitat studies or biomass estimate. Compared to other remote sensing techniques (e.g., SAR or photogrammetry), airborne laser scanning is an adapted tool to provide such information by generating a three-dimensional georeferenced point cloud. Vertical structure analysis consists in detecting the number of layers within a forest stand and their limits. Until now, there is no approach that properly segments the different strata of a forest. In this study, we directly work on the 3D point cloud and we propose a mean shift (MS) based procedure for vertical forest segmentation. The approach that is carried out on complex forest plots improves the discrimination of vegetation strata.

Index Terms— Airborne Laser Scanning, Forest Vertical Structure, Mean shift, Segmentation

1. INTRODUCTION

All plant communities have a vertical structure based on the size and growth pattern of the dominant species. This pattern, called vertical stratification, largely depends on the climatic zone. The main layers are the canopy layer or overstory, the understory, the surface layer or shrub layer, and the ground layer. These layers may have different densities, thicknesses and water contents. Vertical stratification plays an important role in the distribution of fuels, and consequently in fire behaviour, habitat quality, micro-climatic conditions, carbon storage, etc. Airborne Laser Scanning (ALS) is an active remote sensing technology that provides distance measurements between an aircraft (or any other platform) and the surface illuminated by the laser beam. Such range measurements are georeferenced using a GPS/INS system. Depending on the nature of the target, a single pulse emission may be the cause of one to several backscattered echoes. Thereby, ALS is able to penetrate beneath forest canopies down to the ground. Thus, it provides an unstructured 3D point cloud that is a discrete model of the target. Many authors show the potential of multi-echo ALS data to compute Digital Terrain Models (DTM) over vegetated areas, and also to the extract of forest variables [1].

However, emphasis is almost always given to the estimation of tree and crown metrics within the overstory, which do not fully characterize the vertical stratification of heterogeneous forests such as in Mediterranean ecosystems. Only few studies using ALS data focus on the vertical segmentation of forest structures. They mainly

take advantage of the height distribution of the 3D points, and then search threshold values based on histograms by means of different techniques [2, 3, 4, 5]. One can distinguish two approaches: in the plot-based approach, the histogram is calculated over a defined area containing several trees, so that both the mean overstory base height and the mean understory height highly depend on the plot homogeneity. In the tree-based approach, the histogram is defined locally and centred on prior knowledge of the tree positions. This information is thus crucial. Although these approaches work quite well in boreal or managed forests, due to distinguishable gaps between the layers, homogeneous plots and good tree detection, they generally fail in complex environments. The overstory layer can also be bi-storied or multi-storied. Some authors tried to figure out the nature of this layer. To segment trees, the individual 2D tree crown regions detected in each layer of a voxel space are combined [6]. However, this method remains dependent on the correct finding of tree tops used as seed points. A more sophisticated method based on normalized cut segmentation is applied by [3]. However it requires prior knowledge such as the stem position and it is site-dependent since several empirical parameters are involved.

In this paper, we study the potential of the MS algorithm to segment forest. Since its reformulation by [7], the mean shift (MS) has been mainly applied to image segmentation. The MS is a non-linear filter that looks for local maxima (modes) of a density function. It is a non-parametric and unsupervised approach, which only requires a single criterion (the kernel bandwidth). Moreover, the MS can be applied to a joint spatial/spectral space. Here, we explore its properties in the spatial domain. The processing of unstructured ALS point clouds using the MS algorithm was first proposed by [8] who differentiate power lines from vegetation. In this work, we present a new MS-based procedure dedicated to forest stratification. Since the forest point cloud is a multi-modal distribution, the MS technique is used to find the modes that are supposed to be the barycenter of vegetation features. Once it is achieved, the modes are grouped together by height ranges and assigned to vegetation strata.

The MS theory is described in Section 2. Section 3 presents the mathematical bases of the procedure as well as the experimental data set. Section 4 details the results. Finally conclusions are drawn in Section 5.

2. MEAN SHIFT THEORY

The mean shift (MS) is a non-parametric density estimator technique based on the Parzen window kernel. Given n data points $\mathbf{x}_{i=1,\dots,n}$ in a three-dimensional space, the kernel density estimator at point \mathbf{x} can

be written as

$$\widehat{f}_{h,K}(\mathbf{x}) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^n k\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right) \quad (1)$$

with $c_{k,d}$ a normalization constant, h the bandwidth, K the kernel, and $k(\cdot)$ the kernel profile. $k(\cdot)$ describes how strongly the data points are taken into account in the estimation. The MS tries to determine the local modes of the density function $f(\mathbf{x})$, which correspond to the zeros of the gradient $\nabla f(\mathbf{x}) = 0$. Thus, the density gradient estimator can be obtained by

$$\nabla \widehat{f}_{h,K}(\mathbf{x}) = \frac{2}{h^2 c} \widehat{f}_{h,G}(\mathbf{x}) m_{h,G}(\mathbf{x}) \quad (2)$$

with G the kernel profile defined as $g(\mathbf{x}) = -k'(\mathbf{x})$, $c_{g,d}$ a normalization parameter, and $c = c_{g,d}/c_{k,d}$ a normalization constant. In Eq. (2), the first term is the density estimate at \mathbf{x} with kernel G , where

$$\widehat{f}_{h,G}(\mathbf{x}) = \frac{c_{g,d}}{nh^d} \sum_{i=1}^n g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right) \quad (3)$$

and the second term is the so-called MS vector

$$m_{h,G}(\mathbf{x}) = \frac{\sum_{i=1}^n \mathbf{x}_i \cdot g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x} \quad (4)$$

Eq. (4) shows that the MS vector is the difference between the weighted mean (“G-distance”) and \mathbf{x} , the center of the kernel. Eq. (2) can be rewritten as

$$m_{h,G}(\mathbf{x}) = \frac{1}{2} h^2 c \frac{\nabla \widehat{f}_{h,K}(\mathbf{x})}{\widehat{f}_{h,K}(\mathbf{x})} \quad (5)$$

that shows that the MS vector at point \mathbf{x} with kernel G is proportional to the normalized density gradient estimate obtained with kernel K . Thus, it always points towards the direction of the maximum slope of the density function. By setting

$$\mathbf{x}^{t+1} \leftarrow \mathbf{x}^t + m_{h,G}(\mathbf{x}^t) \quad (6)$$

we converge towards the local maxima (t denotes iterations number). Therefore, the MS procedure does not need to evaluate the density function $\widehat{f}_{h,K}$ itself but only the induced kernel $g(\cdot)$. In this study, we use the Epanechnikov kernel,

$$k(\mathbf{x}) = \begin{cases} c_{k,d}(1 - \|\mathbf{x}\|^2) & \text{if inside the unit sphere} \\ 0 & \text{if outside} \end{cases}$$

Then $g(\cdot)$ becomes the indicator function of the unit sphere. In this way, the ratio in Eq. (4) is simply computed as the mean of the points counted within the hyper-sphere of radius h centred on \mathbf{x} . [7] proved that the MS algorithm converges on a stationary point. The MS algorithm can easily be extended to a distance-based segmentation technique, grouping together all the modes that are closer than a certain distance $r \in \mathbb{R}$. Finally, the MS segments are retrieved by aggregation of the basins of attraction of the corresponding convergence points. In accordance with [7], we set $r = h$ for all experiments.

3. METHODOLOGY

3.1. Airborne Laser Scanning Data Set

The ALS data were acquired in June 2008 in north-west Portugal, over a forest dominated by eucalyptus (*Eucalyptus globulus*), in the

framework of a research project financed by the Portuguese Foundation for Science and Technology (FCT). The Riegl LMS-Q560 scanner was run in a full-waveform mode with a pulse repetition of 150 kHz. The average flying altitude of 700 m a.g.l. permitted an average footprint density of 10 pt/m². The point cloud was delivered by the manufacturer after processing the full-waveform data. In order to calculate the height of the objects, the slope effect is removed. For this purpose, the ground points are classified using the TerraScan software [9] and then a Digital Terrain Model (DTM) is generated. Finally, the absolute height of the raw points is calculated. The points classified as ground are kept in the data set and are considered as vegetation hits.

3.2. Adaptive Kernel Bandwidth Selection

Once the Epanechnikov kernel is chosen, the only parameter that needs to be specified is the kernel bandwidth h , *i.e.*, the diameter of the hyper-sphere. This choice is critical because it strongly impacts on the results. A small kernel width leads to more distinct modes (small basins of attraction, more and smaller objects), while a large kernel width aggregates small structures into larger ones (small number of modes with large basins of attraction). The determination of an optimal value of h is actually a major challenge for an efficient MS segmentation.

As far as the vertical component is concerned, the forest layer depth increases with height. Typically, scrubby vegetation strata are thinner than mature tree layers. Thus, in the ALS point cloud, the lower layers are denser than the higher layers. The optimal value of h that allows to distinguish shrubs from understory may fragment a large tree into many segments (lower branches, top foliage, etc.). Fig. 1(a) displays the MS segmentation technique with $h = 1\text{m}$. While the surface vegetation has a coherent shape, the higher features are over-segmented. Increasing h improves the overstory segmentation, but it may cause merging between close small vegetation features. For example, if $h = 4\text{m}$ the dense surface vegetation attracts the sparse understory causing undersegmentation (Fig. 1(b)). Thus, using a unique scale over the entire space is not recommended for the analysis of forest environments, justifying an adaptive kernel bandwidth. Several statistical approaches deal with the scale selection [10]. The kernel bandwidth can also be provided by the user [7], namely for task-dependent applications [11, 12]. Here, is based on the ground forest pattern and we increase h as a function of vegetation height. We establish: $h = 1\text{m}$ as the optimal value if the forest layer is less than 1m ; $h = 2\text{m}$ if its height ranges between 1m and 5 m; finally, we increase the kernel bandwidth to $h = 4\text{m}$ if the forest layer is more than 5 m high.

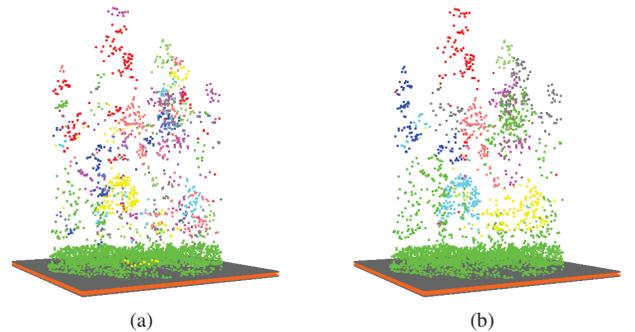


Fig. 1. Mean shift segments using (a) $r = h = 1\text{m}$ and (b) $r = h = 4\text{m}$.

3.3. MS-based Procedure

The procedure is iterative and detection of the strata from the bottom to the top of the forest structure is performed, by adapting the kernel bandwidth within the defined height ranges. First, the 5% height percentile w_0 of the data points $\mathbf{x}_i = (x_i, y_i, z_i)$ is calculated. The adaptive kernel bandwidth is set by its value within the height ranges defined in Section 3.2. Therefore, the modes $\mathbf{x}_i^* = (x_i^*, y_i^*, z_i^*)$ for each \mathbf{x}_i are computed using the corresponding value of h . Then, all modes that are closer than r are grouped together creating MS segments, $C_{p \in \mathbb{N}}$. A forest layer, F_l , is a set of \mathbf{x}_i for which the corresponding MS segments are closer than s from w_0 , with $s/2 = h$. The value of s defines the resolution of the forest stratification, *i.e.*, the number of strata. The ALS assigned points are not taken into account to further calculations. This step improves the segmentation by removing the influence of the lower layers, which are usually denser than the higher. When two regions with different densities are close together, the points of the sparser region are likely to be shifted towards the denser one (Fig. 1(b)). This is more critical when h increases between two iterations. In the second iteration the 5% height percentile of the remaining points is calculated in order to define the new value of h . The procedure keeps on until all points are assigned to one layer. The control parameter equals zero because that guarantees that all points be labeled. If some sparse zones remain unlabeled after the first iterations, the iterative calculation ensures that they are assigned to one layer.

Procedure 1

- 1: **repeat**
- 2: $\forall z_i \in \mathbf{x}_i, \quad w_l = P_{0.05}$
- 3: $\forall \mathbf{x}_i, \quad \mathbf{x}_i^* = \lim_{t \rightarrow \infty} \mathbf{x}_i^t \quad \triangleright$ according to Eq. (6)
- 4: $\forall \mathbf{x}_i, \quad C_p = \left\{ \mathbf{x}_i \mid \forall u = 1, \dots, n \quad \|\mathbf{x}_i^* - \mathbf{x}_u^*\|^2 \leq r^2 \right\}_{p \in \mathbb{N}}$
- 5: $\forall p, \quad F_l = \left\{ \mathbf{x}_i \mid \exists \mathbf{x}_i^* \in C_p \quad \|z_i^* - w_l\|^2 \leq s^2 \right\}_{l \in \mathbb{N}}$
- 6: $\forall \mathbf{x}_i, \quad \mathbf{x}_i = \left\{ \mathbf{x}_i \mid \mathbf{x}_i \notin C_l \right\} \quad \triangleright$ removal of the assigned points
- 7: $\mathbf{x}_i^* = \emptyset \quad \triangleright$ reinitialize modes
- 8: **until** $n = 0$

4. RESULTS AND DISCUSSION

To assess the reliability of the procedure, we use circular plots (200 m²) delimited by systematic sampling. Fig. 2 shows the procedure iterations over a plot with a complex canopy structure. It is clear that the original points progressively converge producing coherent segments by means of the MS algorithm, at least along the vertical component. It can be visually assessed that they are well assigned to layers. This proves that the MS-based procedure is reliable to find the center of mass of vegetation layers. Since these barycenters are represented by the MS modes, our procedure succeeds in stratifying the forest.

A distinctive characteristic of our procedure is the removal of the assigned points during the iterations for the further calculations. By increasing h , the MS vector is computed within a large neighbourhood. When two layers are very close, the boundary points from the sparser layer are shifted towards the denser layer. This phenomenon that may occur between the understory and the overstory is more critical in the transition between surface vegetation and the understory. A small kernel bandwidth ($h=1$ m) better applies to

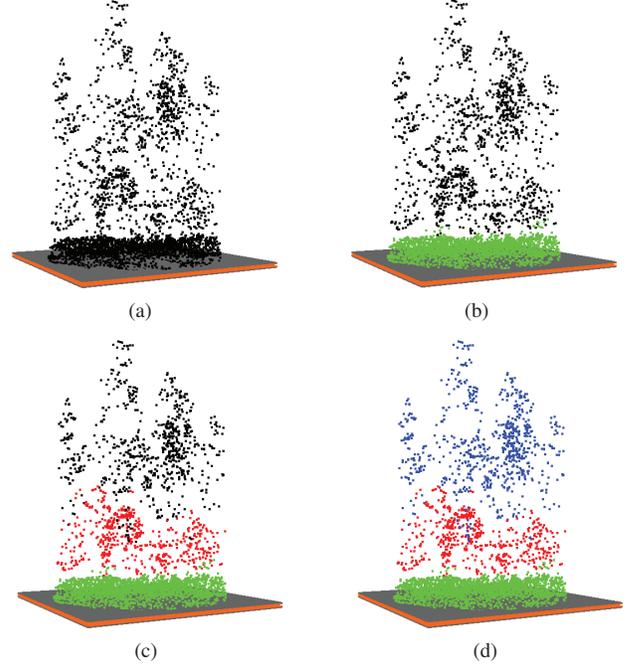


Fig. 2. Procedure iterations using the parameters defined in Section 3.2. (a) Original ALS points (black). (b) 1st iteration, points assigned to first layer (green) with $w_1=0$ m and therefore $s/2=r=h=1$ m. (c) 2nd iteration, points assigned to second layer (red) with $w_2=3$ m, thus $s/2=r=h=2$ m. (d) 3rd iteration, points assigned to third layer (blue) with $w_2=8$ m, thus $s/2=r=h=4$ m.

dense surface vegetation. When h is increased to segment understory features, step 6 of the procedure allows these vegetation hits to converge, avoiding the influence of near denser zones. The more complex the forest structure, the more meaningful this step.

Up to now, separation of the forest layers was performed using 3D point cloud derived products by means of different methods. Our procedure takes advantage of the mode space, where the layers are more likely to be identified. An advanced analysis of the segmentation between the understory and the overstory confirms it. One can notice that our procedure leads to the same result (Fig. 2) when the upper height threshold moves within a 4 m range. When h increases from 2 m to 4 m (Section 3.2), the upper height threshold w may be defined from 4 m until 8 m. This is definitely an improvement in forest stratification. In complex forest structures where point cloud based methods fail, our procedure allows height thresholds to be defined in a broader range. This is because forest stratification is performed on the mode space where the ALS points are shifted towards (down or up) the center of mass of the vegetation strata creating a gap between the layers.

When increasing the resolution, *i.e.*, keeping $h=2$ m until $w \leq 10$ m, a fourth layer is detected (Fig. 3) which suggests a layer of dominated trees. This shows the potential of the procedure to differentiate the main layers into single-storied and bi-storied. However, setting the w value for these thinner resolutions is tricky, because keeping $h=2$ m to higher elevations may split a single tree into two or more layers. The post-processing of each layer derived from a coarser analysis is a field to investigate. This analysis also shows that setting

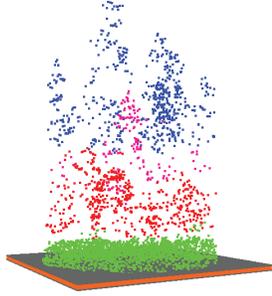


Fig. 3. Result of the procedure using $h=1\text{m}$ if $w_l \leq 1\text{m}$, $h=2\text{m}$ if $1\text{m} < w_l \leq 10\text{m}$, and $h=4\text{m}$ if $w_l > 10\text{m}$. Compared with Fig. 2(d), one more layer is detected (violet) within the points assigned to the upper layer (blue).

three thresholds does not necessarily result in three layers. The number of layers is actually inherent to the forest structure. To assess the robustness of the procedure, a larger plot (2500 m^2) with a heterogeneous structure is tested. The data are processed in two ways. First, we analyse the entire plot. Second, the plot is divided into four parts of equal area, then processed independently, and finally the results reassembled. The parameters are those defined in Section 3.2. The results are very similar. Fig. 4 illustrates the second method. The good separation between the three main layers is clear. Fig. 4 also shows that the procedure applies to various forest realities, from simple to complex structures. Moreover, it does not split the layers horizontally. Vegetation structures are well represented and their ranges can be locally calculated. Additionally, the procedure works well within very complex and heterogeneous areas where the layers are sometimes mixed, *i.e.*, the top of the understory is higher than the bottom of the overstory within a neighbourhood.

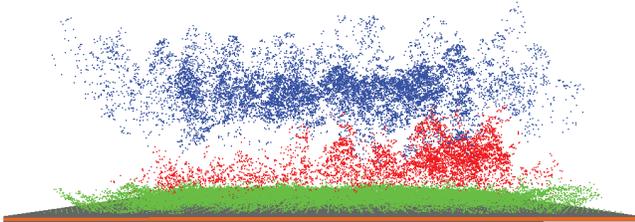


Fig. 4. Forest layers resulting from the procedure using the height ranges defined in Section 3.2. The forest plot (2500 m^2) is stratified into surface vegetation (green), understory (red), and overstory (blue).

5. CONCLUSIONS

In this work, we introduce the MS potential to segment vegetation features working directly on the ALS 3D point cloud. Unlike other methods, forest segmentation is carried out in three dimensions providing genuine 3D segments. Additionally, a procedure is proposed to detect forest layers. We simplify such structures using the MS algorithm and then segment the forest layers in the mode space. We show that the proposed procedure is very valuable to analyse complex forest structures. In order to improve its robust-

ness, more efforts must be put into the adaptive kernel bandwidth settings. Studying the mode height distributions should improve the procedure, by making it more independent of the forest type and the point cloud characteristics. Since the MS applies to a joint space (spatial and attributes), additional attributes of the ALS points (intensity, number of echoes, etc.) may be introduced in the MS vector calculation. The analysis of different kernel functions may not only help to diminish the dependency of the procedure on the input parameters, but also to better fit the tree crowns aiming at single tree extraction. In the near future, we pretend to validate our procedure using forest inventory data acquired in the framework of the above referred FCT project.

6. REFERENCES

- [1] J. Hyypä, H. Hyypä, D. Leckie, F. Gougeon, X. Yu, and M. Maltamo, "Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests," *International Journal of Remote Sensing*, vol. 29(5), pp. 1339–1366, 2008.
- [2] D. Riano, E. Meier, B. Allgöwer, E. Chuvieco, and S. Ustin, "Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling," *Remote Sensing of Environment*, vol. 86(2), pp. 177–186, 2003.
- [3] J. Reitberger, C. Schnörr, P. Krzystek, and U. Stilla, "3D segmentation of single trees exploiting full waveform LIDAR data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64(6), pp. 561–574, 2009.
- [4] A. Barilotti, F. Sepic, and E. Abramo, "Automatic detection of dominated vegetation under canopy using Airborne Laser Scanning data," *SilviLaser (Edinburgh, UK)*, pp. 134–143, 17–19 September 2008.
- [5] S. Popescu and Z. Kaiguang, "A voxel-based lidar method for estimating crown base height for deciduous and pine trees," *Remote Sensing of Environment*, vol. 112(3), pp. 767–781, 2008.
- [6] Y. Wang, H. Weinacker, B. Koch, and K. Sterenczak, "Lidar point cloud based fully automatic 3D single tree modelling in forest and evaluations of the procedure," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37 (Part B6b), pp. 45–51, 2008.
- [7] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24(5), pp. 603–619, May 2002.
- [8] T. Melzer, "Non-parametric segmentation of als point clouds using mean shift," *Journal of Applied Geodesy*, vol. 1(3), pp. 159–170, 2007.
- [9] A. Soininen, "Terra Scan for Microstation," *User's guide*, 1999.
- [10] D. Comaniciu, V. Ramesh, and P. Meer, "The variable bandwidth Mean Shift and data-drive scale selection," *IEEE International Conference on Computer Vision*, vol. 1, pp. 438–445, 2001.
- [11] S. Bo, L. Ding, H. Li, F. Di, and C. Zhu, "Mean shift-based clustering analysis of multispectral remote sensing imagery," *International Journal of Remote Sensing*, vol. 30(4), pp. 817–827, 2009.
- [12] X. Huang and L. Zhang, "An adaptive mean-shift analyses approach for object extraction and classification from urban hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46(12), pp. 4173–4185, 2008.