Tree filtering for high density airborne LiDAR data

M.Z.A. Rahman, B. Gorte

Optical and Laser Remote Sensing, Department of Earth Observation and Satellite System (DEOS),
Faculty of Aerospace Engineering, Delft University of Technology, the Netherlands–
M.Z.AbdRahman@tudelft.nl, B.Gorte@tudelft.nl

Abstract

A high resolution Airborne LiDAR data creates better opportunity for an individual tree measurement and provides valuable results for more precise forest inventory. This paper presents tree filtering approach that able to separate dominant tree and undergrowth vegetation. The results can be used for a detailed individual tree measurement. This process is one of the main steps for a single tree extraction from the high resolution Airborne LiDAR data. The filtering technique lies on the fact that a dominant tree has distinct parts in the histogram that represent tree crown, tree trunk, and ground surface with or without undergrowth vegetation. The shape of the histogram was used to identify points that belong to the tree crown and the tree trunk. More points were assigned to the tree trunk based on an iterative analysis of the histogram at certain height above the ground surface. This step was coupled with the RG segmentation. It was found that the filtering routine failed to remove very close undergrowth vegetation. It was also observed that in order to get a good result, the tree filtering method needs at least small area of the tree trunk.

Keywords: *High resolution Airborne LiDAR, RG segmentation, 1D Gaussian filter, Gaussian fitting*

1. Introduction

Laser scanning is now becoming one of the important sources of information for forest applications. The laser beam with specific settings may be able to penetrate the forest structure, thus giving a better opportunity for accurate forest variable measurements. Hyyppä et al. (2004) has listed out numerous techniques and algorithms for tree variable extraction. The features and the predictors in the statistical method are being assessed from the laser derived surface models and point clouds. This information is then used to estimate forest parameters based on regression and discriminant analysis (Table 1). On the other hand, the image processing methods use the neighborhood information of point clouds and pixel of a Digital Surface Model (DSM). The physical features such as, tree crowns, individual trees, group of trees or the whole stands can be derived using this method. In this method, further step of forest parameters extraction are assessed using the existing models and statistical methods.

Table 1 : Tree variable extraction based on statistical method of LiDAR data (Hyyppä, et al., 2004)

Method	Description on method	Forest properties
Canopy profile area	The *canopy profile area* is directly related to the logarithm of the timber volume	Volume of timber
Height percentiles of the distribution of canopy heights	The *Height percentiles of the distribution of canopy heights* as predictors in regressions models to estimate mean tree height, basal area and volume	Predictors in regressions models to estimate mean tree height, basal area and volume
Canopy reflection sum, ground reflection sum and Canopy closure	*Canopy reflection sum* is the sum of the portion of the waveform return reflected from the canopy. *Ground reflection sum* is the sum of waveform return reflected from the ground multiplied by a factor correcting the canopy attenuation. *Canopy closure* is approximated by dividing the sum of the canopy and ground reflection sums	Predictors in regressions models to estimate tree height, basal area and volume
Canopy height and density metrics	*Canopy height metrics* included e.g. quantiles corresponding to the 0,10,…,90 percentiles of the first pulse laser canopy heights and corresponding statistics, where as *canopy density* corresponded to the proportions of both first and last pulse laser hits above the 0,10,…,90 quantiles to total number of pulses	Canopy height and density metrics
Tree cover and Surface cover	*Tree cover* is calculated from the proportion of laser hits from tree canopy divided by the total number of laser hits. *Surface cover* is defined as the proportion of laser hits from the surface and the total number of hits	Area of the tree and area of the ground surface
Relative standard deviation of tree heights, the proportion of single returns and the proportion of first return, proportion vegetation points, mean intensity, standard deviation of both single and surface returns	*The proportion vegetation point* is defined as a number of returns that are located above the crown base height divided by the total number of returns from the segment. This information is used for tree species classification	Tree species classification
Crown shape	*Crown shape* is defined by fitting a parabolic surface to the laser point cloud	Crown shape

Litkey et al.(2007) pointed out that there are two main feature extraction methods that can be used to derive forest information from Airborne LiDAR data. The first method is based on a statistical canopy height distribution (e.g., Naesset (1997)) and the second approach is based on an individual tree detection (e.g., Hyyppä and Inkinen (1999) and Persson, et al. (2002)). It was stated that the methods based on the statistical canopy height distribution typically use regression, non-parametric or discriminant analysis for forest parameter estimation. On the other hand, the individual-tree-based method uses the neighbourhood information of canopy height point clouds and the pixels of Canopy Height Model (CHM) to extract features such as crown size, individual tree height and tree location. The forest inventory data are then being estimated using existing models and statistical techniques.

Numerous studies stated that a discrete return laser scanner data can produce accurate information on a tree canopy since the quantiles of height distribution of laser scanner data area related to the vertical structure of the tree canopy (Maltamo, et al., 2004). Furthermore, since some of the laser pulses penetrate the canopy of dominant tress, it is possible to analyze undergrowth vegetation. In their study, Maltamo et al. (2004) used a histogram plot to analyze multi-layered canopy structure. In this study it was concluded that the characteristics of the canopy height laser point data, especially the shape of the height distribution can be used to identify multi-layered stand structures. Reitberger et al. (2007) introduced a method to delineate tree crown and detection of stem position of single a trees from dense Airborne LiDAR data. In this study, trees were delineated using a watershed algorithm on the CHM and the possible stem position was derived from the local maxima of the CHM. In this study, they have introduced a 3-step algorithm to search stem position in each tree segment. Firstly, all the points between the ground and the crown base height were separated and the points were clustered using hierarchical clustering based on their horizontal distances. Finally, the stem position was estimated using a robust RANSAC-based adjustment of the stem points.

The objective of this study is to develop a new tree filtering approach for high density airborne LiDAR data that is able to separate dominant trees and undergrowth vegetation. The filtering process is one of the main steps of individual tree variable measurement (refer Figure 1). In this paper, the filtering method was tested on different LiDAR datasets with different density of undergrowth vegetation. The results can be used for individual tree variable measurements of dominant trees and undergrowth vegetation. In this case, the tree measurement can be carried out directly on a single tree rather than based on the regression models.

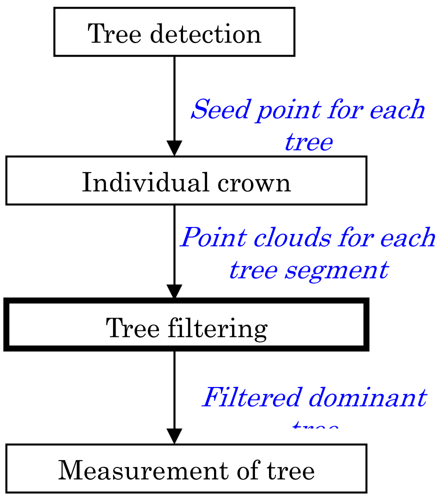


Figure 1: The overall flow for an individual tree measurement

2. Materials and method

2.1 Study site

This study was conducted at the Duursche Waarden floodplain, the Netherlands. The floodplain is along the IJssel River, which is the smallest tributary of the Rhine River in the Netherlands (Straatsma and Middelkoop, 2006). This area is partly covered by meadow and arable land and most of the areas have become nature. The vegetation in this area comprises of (1) softwood forest Willow (Salix abla, Salix viminalis), poplar (Populus nigra, Populus x canadensis), (2) hardwood forest oak (Quercus robur), ash (Fraxinus excelsior) and a small pine stand (Pinus sylvestris) on a river dune, together with (3) reed marshes (Phragmites australis), and (4) herbaceous vegetation with sedge (Carex hirta), sorrel (Rumex obtusifolius), nettle (Urtica dioica), thistle (Crisium arvense) and clover (Trifolium repens).

2.2 LiDAR data

The LiDAR data used in this study was captured by the FLI-MAP 400 system. The FLI-MAP 400 is a helicopter mounted LiDAR system designed to capture highly detailed terrain features with high accuracy. It was claimed that the absolute accuracy of the FLI-MAP 400 data measured over hard and level surfaces is 2.5 to 3.0 cm. The system is capable of scanning in three directions (forward, down (nadir) and back) and this increases the chance of capturing a significant amount of reflected pulses from the ground even in a quite densely vegetated area. The FLI-MAP 400 data records maximum four laser reflections with an unmatched distance of 0.9 m, which enables optimal interpretation of a detailed terrain model even in vegetated areas. The data with an average density of 70 points per meter square were acquired during winter in 2007. The leaf-off data allow better penetration through a tree canopy and therefore the vertical structure of a tree can be easily revealed. In this study, 10 sample trees were selected with different tree species and undergrowth density. All samples were delineated manually and for further processing stage, each sample was attached with one seed point located on top of the tree.

2.3 Histogram-based tree filtering

In this study, the new tree filtering approach is called a histogram-based tree. This method relies on the fact that a dominant tree would have distinct parts in the histogram that represent tree crown, trunk, ground surface and undergrowth vegetation. Previous study by Straatsma and Middelkoop (2006) has shown that the shape of height distribution of a tree has a higher frequency of laser pulses from the crown and undergrowth vegetation. On the other hand, the reflected laser pulse from the trunk is at a lower frequency. The segmentation process starts from a seed point located on top of the tree crown, and the shape of histogram is used to identify points that belong to the tree crown and the tree trunk. The RG segmentation is then used to subdivide the points into the tree crown and the trunk. The search for the tree trunk continues by iteratively analyses the shape of the histogram at certain height above the ground. This process is coupled with the RG segmentation to assign additional points to the tree trunk. The process continues until it is no longer able to distinguish between tree trunk and the undergrowth vegetation. Furthermore, if the process stops before it reaches the ground surface, the tree trunk is extrapolated by fitting a three dimensional line (3D line) using the points which have been previously assigned as a tree trunk. The additional points for a tree trunk is then collected based on the distance between the line and the remaining point clouds.

2.5 One-dimensional (1D) Gaussian filtering

As explained earlier in section 2.3, the histogram of the point cloud distribution of a single tree was used as a reference to assign points into tree crown, tree trunk, undergrowth vegetation and ground surface. In this study, the boundary that marks each part of the tree on the histogram was defined automatically using a multi-modal Gaussian fitting routine. It was observed that, the original histogram contains noises that need to be removed in order to get better result in Gaussian fitting process. Thus, the first step was to smooth the histogram. A 1D Gaussian filter was used to smooth out the histogram surface. In this study, only one value of sigma (0.015) of Gaussian filter was used for all the datasets (refer Figure 2).

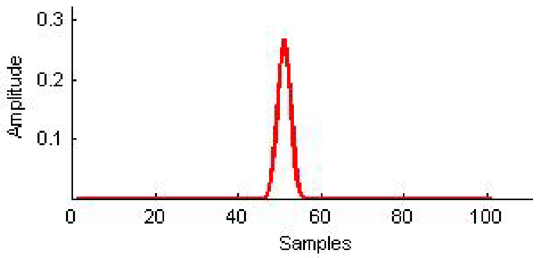
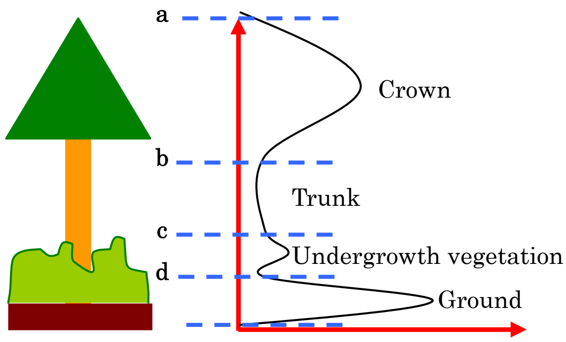


Figure 2: One-dimensional Gaussian filter

2.6 Gaussian fitting on histogram

A Gaussian fitting on the histogram was based on nonlinear curve-fitting problems in a least square sense which is available in Matlab (lsqcurvefit). This routine determines the possible number of Gaussian peaks based on the pre-defined values such as number of possible Gaussian shapes and Gaussian model parameters (sigma, position, frequency). In order to determine this information, the peaks in the histogram of a single tree can be assumed to have a composition of tree crown, undergrowth vegetation and ground surface (refer Figure 3). As depicted in Figure 3, the Gaussian fitting routine was then applied on the filtered histogram to define a specific boundary for these 3 parts. Each boundary was defined by 3σ value from the Gaussian each peak (μ).



a - Starting level (elevation) for tree crown
b - Starting level (elevation) for tree trunk
c - Starting level (elevation) for undergrowth vegetation
d - Starting level (elevation) for ground

Figure 3: Shape of histogram for a single tree

All tree samples were delineated manually by hand and a seed point was attached on top of each tree. A semi-automatic tree detection and crown segmentation will be explained later in another study. The histogram-based tree filtering process was carried out with the following steps:

1. Place a seed point on top of each tree.
2. Define growing distances for 3 parts, 1) tree crown, 2) tree trunk and 3) distance between a 3D line and point clouds to extract additional points for tree trunk
3. Calculate a histogram for a single tree and filter the histogram with 1D Gaussian filter
4. Fit a Gaussian function on the filtered histogram to extract 3 different parts of the tree, namely, 1) tree crown, 2) undergrowth vegetation and 3) ground surface
5. RG segmentation from the tree crown to the level that marks the beginning of the tree

trunk
6. RG segmentation for the tree trunk
7. Iteratively analyze the shape of the histogram to add more points to the tree trunk
8. Stop step (7) if the process is no longer able to distinguish between points that belong to the tree trunk, the undergrowth vegetation as well as the ground surface
9. Create a 3D line based on the points that have been classified as a tree trunk
10. Assign additional points to the tree trunk based on their distances to the 3D line.

Step (9) creates a 3D line, which intends to extrapolate tree the trunk until it reaches the ground surface. This will be the last step of collecting points for the tree trunk, since the filtering process as indicated in step (8) was no longer able to distinguish between the points that belong to the tree trunk, the undergrowth vegetation and the ground. This step assigned more points to the tree trunk by selecting points at certain distance from the extrapolated tree trunk (3D line). The tree filtering method basically needs three input parameters, namely growing distance for the tree crown, growing distance for the tree trunk, and distance between points to the interpolated 3D line. In general, large growing distance value was used for segmenting the tree crown, and small growing distance value was used for the tree trunk instead.

3. Results and discussions

The results showed that the histogram-based method performs quite well in separating the dominant trees and the undergrowth vegetation (refer Figure 5). Furthermore the 1D Gaussian filtering helps in reducing noises in the original histogram and enhanced the general shape of the histogram. This process subsequently eased the multi-modal Gaussian fitting on the smoothed histogram (refer Figure 4). However, it should be noted that, different setting of the Gaussian filter for example with different σ value would produce different result on a smoothed histogram. This could suppresses some useful information and reduce the effectiveness of the tree filtering method.

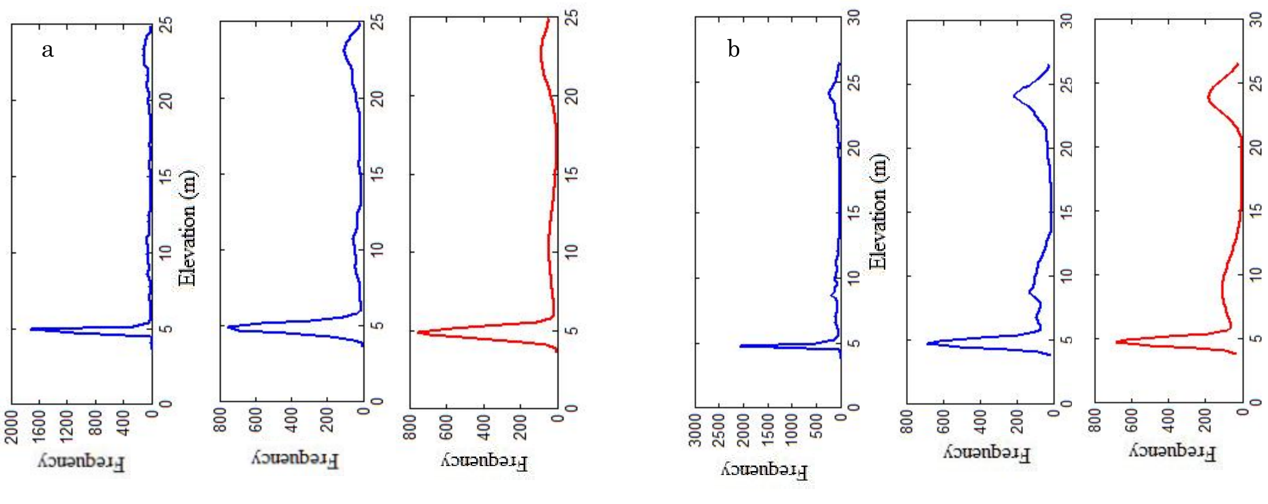


Figure 4: Examples of the original histogram, filtered histogram and fitted histogram for tree 1 (a) and tree 2 (b)

Figure 4 shows some examples of the original histogram, the smoothed histogram and the estimated Gaussian functions on the histogram.

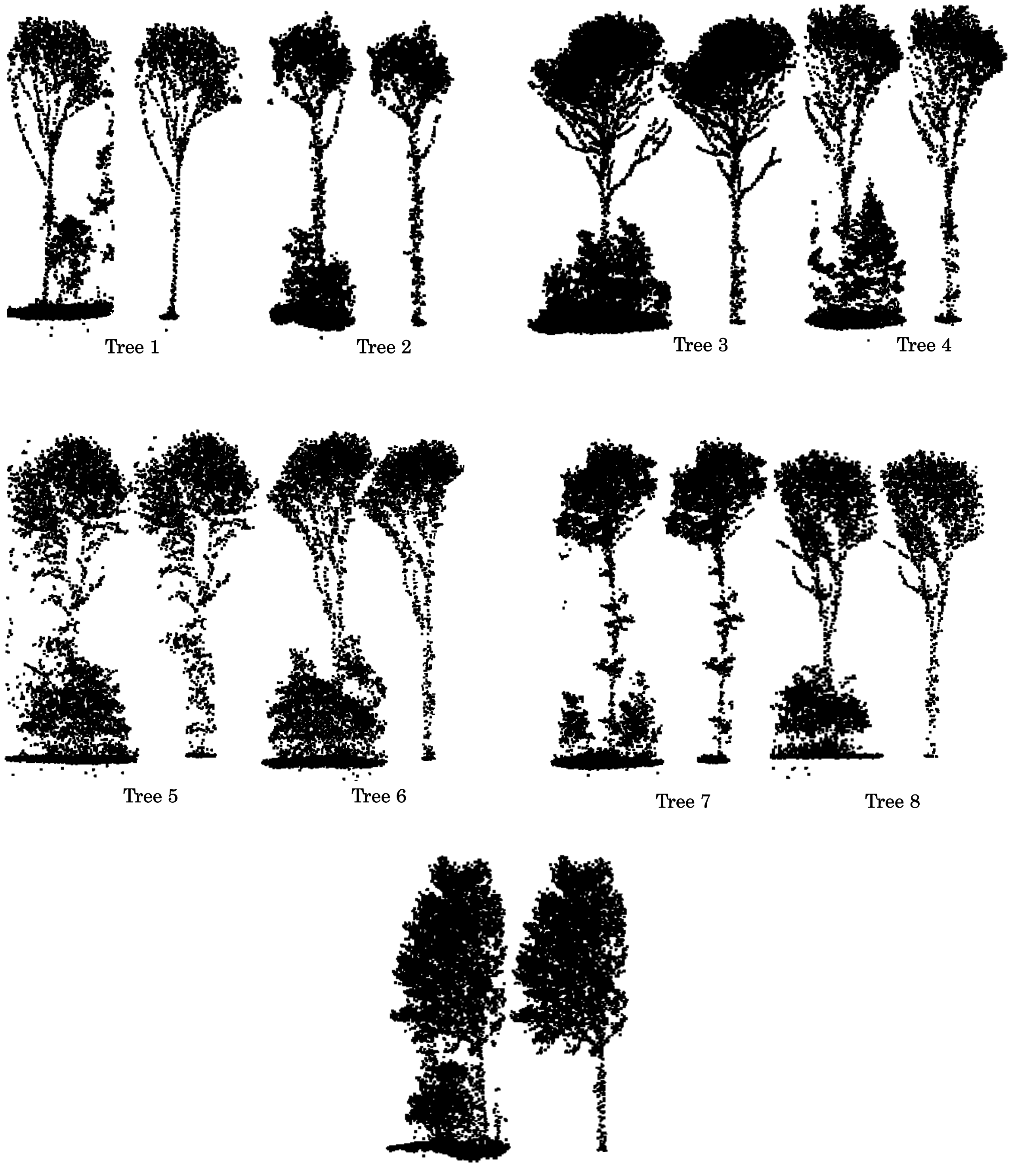


Figure 5: Original trees and filtered trees

In this study it was found that the histogram-based tree filtering method requires at least small area of a tree trunk and the reflected laser pulses from this area should be less than the tree crown. In this case it would be rather difficult for trees with dense branches along the tree trunk. Very small area of tree trunk caused overlapping boundary between the tree crown and the tree trunk. Thus, 3σ value will not be appropriate to represent the boundary of each part. Figure 7 (a) shows an example for a tree condition where there is a very small area of tree trunk and the undergrowth vegetation is very close to the dominant tree. In this example, a special experiment was conducted to observe the size of the area for the tree trunk in the histogram. For this purpose, the σ value for 1D Gaussian filter was tuned from 0.0026 to 0.1 and the different between two levels (between b and c) was observed (refer figure 3). It was found that the Gaussian fitting routine failed to identify appropriate value for level b and c, in which the different between them (level b – level c) should have a positive value. Figure 7 (b) shows that the histogram-based approach failed to separate the dominant tree and the undergrowth vegetation.

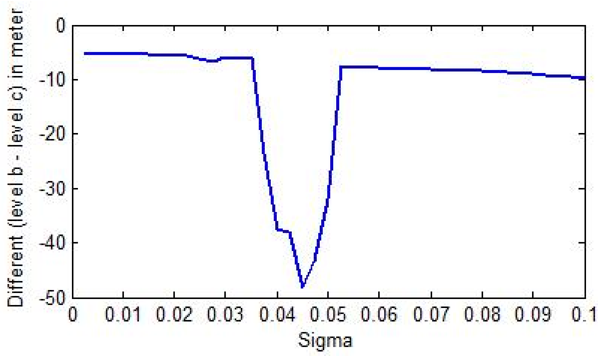


Figure 6: The different in meter for level b and level c

In this study, it was also observed that each tree requires different value of growing distance depends on the closeness of the undergrowth vegetation to the tree crown and the tree trunk (refer Table 2). Small growing distance should be used for very close undergrowth vegetation. Therefore, further study is required to optimize the tree filtering method, in which values for the growing distance should be defined based on the density of the undergrowth vegetation.

Table 2: Growing distance for each tree

Dataset	**Growing distance for tree crown (m)**	**Growing distance for tree trunk (m)**	**Growing distance for 3D line (m)**
Tree 1	0.8	0.5	0.6
Tree 2	0.8	0.6	0.5
Tree 3	0.8	0.4	0.5
Tree 4	0.8	0.6	1.0
Tree 5	0.8	0.6	1.0
Tree 6	0.5	0.4	0.5
Tree 7	0.8	0.6	1.1
Tree 8	0.5	0.4	0.5
Tree 9	0.5	0.3	0.5

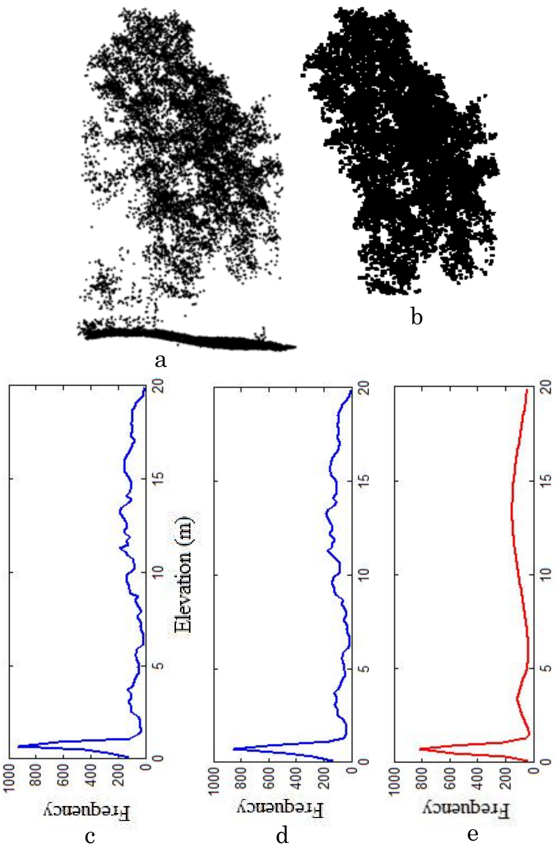


Figure 7: Original tree (a), filtered tree (b), original histogram (c), filtered histogram (d) and fitted histogram (e)

4. Conclusions

In general the histogram-based tree filtering method which aims at separating the dominant tree and undergrowth vegetation performed well on all datasets. The results can be used in further detailed tree variable measurement for instance, species identification, stem diameter, crown size, crown volume and etc. However, the filtering method failed to filter the dominant tree which is very close to the undergrowth vegetation. It was also shown that the filtering method still needs to be optimized by taking into account the density of the undergrowth vegetation. This information will be used as the basis to select proper growing distance values for tree crown, tree trunk and 3D line. Further study is also required to quantify the effect of different magnitude (σ) of the 1D Gaussian filter to the performance of the histogram-based tree filtering method. In future, this method will be applied together with the tree detection and crown delineation routines on a larger Airborne LiDAR dataset.

Acknowledgments

The authors would like to thank Rijkwaterstraat, The Netherlands for giving an opportunity to use the FLI-MAP 400 Airborne LiDAR data.

References

Hyyppä, J., et al. 2004. Algorithms and Methods of Airborne Laser-Scanning for Forest

Measurements. In:. M. Theis, B. Koch, H. Spiecker, H. Weinacker (Eds.). *Proceedings of ISPRS working group VIII/2 "Laser-Scanners for Forest and Landscape Assessment"*. Freiburg, University of Friburg: 82-89

Hyyppä, J., and M. Inkinen., 1999. Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*, 16, 27-42.

Litkey, P., et al. 2007. Waveform features for tree identification. In: *Proceedings of ISPRS Workshop on Laser Scanning 2007 and SilviLaser.* Espoo, Finland: 258-263.

Maltamo, M., et al., 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sensing of Environment*, 90, 319-330.

Naesset, E., 1997. Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing,* 52, 49-56.

Persson, A., J. Holmgren, and U. Soderman., 2002. Detecting and Measuring Individual Trees Using an Airborne Laser Scanner. *Photogrammetric Engineering Remote Sensing*, 68, 925 - 932.

Reitberger, J., et al., 2007. Single tree detection in forest areas with high-density LiDAR data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* 36, 139-144.

Straatsma, M. W., and H. Middelkoop., 2006. Airborne Laser Scanning as a Tool for Lowland Floodplain Vegetation Monitoring. *Hydrobiologia,* 565, 87-103.