

Article

Correction of Erroneous LiDAR Measurements in Artificial Forest Canopy Experimental Setups

Renato Cifuentes ^{1,*}, Dimitry Van der Zande ², Christian Salas ³, Jamshid Farifteh ¹ and Pol Coppin ¹

¹ Department of Biosystems, Katholieke Universiteit Leuven, M3-BIORES, Willem de Croylaan 34, BE-3001 Leuven, Belgium; E-Mails: jamshid.farifteh@biw.kuleuven.be (J.F.); pol.coppin@biw.kuleuven.be (P.C.)

² Management Unit of the North Sea Mathematical Models, Gulledele 100, BE-1200 Brussels, Belgium; E-Mail: dimitry.vanderzande@mumm.ac.be

³ Departamento de Ciencias Forestales, Universidad de La Frontera, Temuco PO-Box 54-D, Chile; E-Mail: christian.salas@ufrontera.cl

* Author to whom correspondence should be addressed; E-Mail: renato.cifuentes@biw.kuleuven.be; Tel.: +32-16-329750; Fax: +32-16-322966.

Received: 30 December 2013; in revised form: 28 May 2014 / Accepted: 10 June 2014 /

Published: 3 July 2014

Abstract: Terrestrial laser scanning (TLS) data makes possible to directly characterize the three-dimensional (3D) distribution of canopy foliage elements. The scanned edges of these elements may result in incorrectly point measurements (*i.e.*, “ghost points”) impacting the quality of point cloud data. Therefore, estimation of the ghost points’ spatial visibilities, measurement of their characteristics and their removal are essential. In order to quantify the improvements on data quality, a method is developed in this study to efficiently correct for ghost points. Since the occurrence of ghost points is governed by a number of factors, (e.g., scanning resolution and distance, object properties, incident angle); the developed method is based on the analysis of the effects of these factors under controlled conditions where canopy-like objects (*i.e.*, leaves, branches and layers of leaves) were scanned using a continuous-wave TLS system that employs phase-shift technology. Manual extraction of ghost points was done in order to calculate the relative amount of ghost points per scan, or ghost points ratio (*gpr*). The *gpr* values were computed in order to: (i) analyze their relationships with variables representing the above factors; and (ii) be used as a reference to evaluate the performance of filters used for extraction of ghost points. The ghost points’ occurrence was modeled by fitting regression models using different predictor variables

that represent the variables under study. The obtained results indicated that reduced models with three predictors were suitable for *gpr* estimation in artificial leaves and in artificial branches, with a relative root mean squared error (RMSE) of 4.7% and 3.7%, respectively; while the full model with four predictors was appropriate for artificial layers of leaves, with relative RMSE of 4.5%. According to the statistical analysis, scanning distance was identified as the most important variable for modeling ghost points occurrence. Results indicated that optimized distance-based filters relative to the scanning distance have improved the outcomes in ghost points detection, in comparison to standard filtering criteria. These results suggest that more accurate characterization of forest canopy 3D structures can be achieved by removing ghost points using the new developed method.

Keywords: continuous-wave LiDAR; mixed pixel effect; ghost points filter; experimental setup

1. Introduction

Quantifying the three-dimensional (3D) structure of an individual tree or forest stand is a challenging task due to the complex organization of all the elements involved, predominantly at the canopy level (e.g., stems, branches, and leaves) [1]. Diversity exists in the shape and size of these elements, in the range of plant form and in the combination of forms that give a canopy architecture [2–4]. Light detection and ranging (LiDAR) remote sensing used for terrestrial laser scanning (TLS), is a tool to generate point clouds (x, y, z) that describe canopy structure as a whole, providing information about the distribution of biomass [5], and also characterizes the spatial distribution of canopy elements at a high level of detail [6]. Three primary types of TLS technologies are being employed in commercial laser scanners [7]: (i) time-of-flight discrete-return scanners; (ii) time-of-flight waveform scanners; and (iii) continuous-wave phase-shift scanners. Time-of-flight discrete-return instruments provide high accuracy at large range. This category of scanners has been the most used for vegetation structure assessments [8]. In addition to the point cloud data, most TLS instruments also record the instrumental (*i.e.*, raw) point intensity value. It is well known that intrinsic errors in TLS-based range measurements as well as errors due to the interaction of the laser beam with the environment impact the quality of these point clouds. As reported by [9], extra noise might be added due to the reflective properties of objects and angle of incidence of the laser beam. Similarly, regarding the laser beam/object interaction, an important source of noise is the imperfect scanning of edges of the measured elements. This condition, defined as a “mixed pixel” effect [10], occurs at spatial discontinuities in the space wherever the laser beam lies partially on two surfaces with different distances from the sensor. In this scenario, the laser beam is reflected by both the foreground and background surfaces, and the sensor receives a mixture of the two signals. Depending on the type of laser scanner, the resulting range measurement may be reported at the distance to the foreground object, the background, somewhere in between, or, even at distances closer than the foreground object or further than the background [11–13]. The incorrectly measured points affecting the subsequent processing of these data [14] will be called hereafter “ghost points”. Scenes with a high number of edges or discontinuities, in color or distance,

are more prone to the formation of ghost points. The proportion of ghost points may increase significantly when the scanned areas include dense understory and/or thin objects.

Considering the measurement setup, a number of variables govern ghost points occurrence, namely, scanning resolution and distance, object properties (e.g., size and reflectivity) and incident angle [12]. From these variables, only the scanning resolution is internal to the scanner and user-configurable, while the rest are external to the scanner. The effect of these variables on the occurrence of ghost points have been comprehensively studied and discussed in [9,12,15–18]. Previous research has focused on understanding the sources of error in the sensing and modeling process [18–20]. Similarly, other studies have aimed to characterize laser instruments and analyze the effect of various operating parameters in order to identify edges and remove the unwanted data points by means of e.g., two-dimensional edge-detection processes [21], and algorithms to detect depth discontinuity and mixed pixels in 3D data [20,22]. A manual selection and correction of the point cloud during the TLS data preprocessing procedure has also been tested [9], but this technique is not efficient to delete ghost points on large datasets (e.g., real forest canopies). More recent studies, however, have improved quality of TLS data collected in forest environments by investigating the morphology of canopy gaps and correcting for the bias introduced by points on time-of-flight TLS data [23]. Improved data is also obtained when default and customized filters (*i.e.*, threshold filters for range and intensity) are used to remove ghost points on datasets from continuous-wave and time-of-flight TLS systems [7]. However, these filters do not provide consistent improvement throughout the whole dataset, mainly due to the limitations of TLS systems (e.g., effective range, radial scanning) and their specific characteristics (e.g., beam diameter and divergence) [12]. In order to accurately develop filtering techniques for removing ghost points, it is essential to understand which variables, internal and/or external to the instrument, contribute the most to their occurrence. This is especially important in forest canopy research, since ghost points represent true interceptions of the laser beam, therefore, they should be considered in gap probability analysis (e.g., leaf area index, canopy cover), but have to be filtered out for structural analysis (e.g., study of single stem shape) [7]. In addition, since sensitivity of edge detection algorithms decreases with increasing range, more accurate results can be obtained by selecting adaptive threshold functions, so that within the ranges of interest the filter could be as sensitive as possible [21].

The main goal of this paper is to quantify the improvements on data quality by measuring the characteristics of the ghost points obtained from TLS measurements. This was done under controlled conditions, using artificial measurement setups, simulating forest canopy scenes and applying adaptive techniques for detection and correction of ghost points. The second objective is to estimate the effects of object properties, scanning resolution, distance and incident angle on the quality of continuous-wave TLS data, in terms of ghost points occurrence.

2. Materials and Methods

2.1. TLS Instrument

The TLS instrument used in this study was the panoramic-type FARO[®] LS 880HE (FARO[®] Technologies Inc., Stuttgart, Germany), which uses a continuous-wave laser operating at 785 nm and

the phase-difference technique to measure the 3D position of objects within a range of 76 m. The laser power of the scanner is 20 mW and the returning intensity is recorded in 11 bits (0 to 2048). While operating, the laser beam is deflected by a rotating mirror, giving the angular coverage of 320° on the vertical plane. A rotating motor allows the 360° azimuth scan. For the present study, the scanning area was reduced to the area of interest, in order to avoid unnecessary data recording.

2.2. Measurement Setups

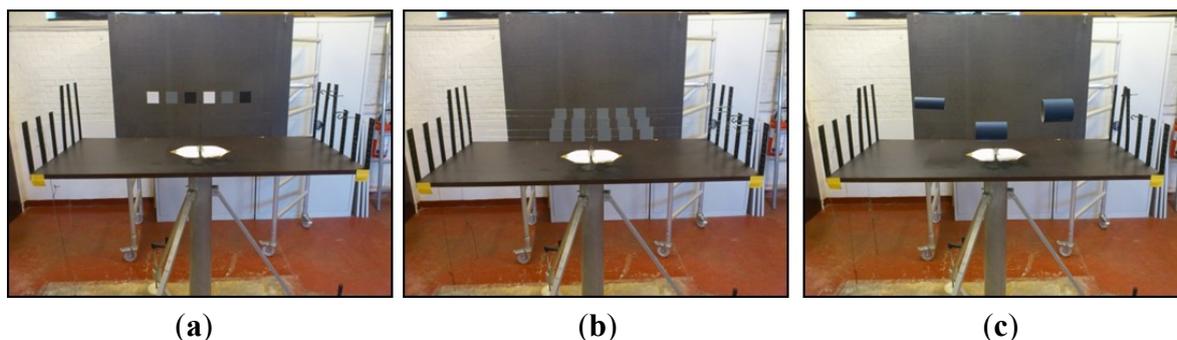
The scanning distance, incident angle, and scanning resolution were modified throughout a series of controlled indoor TLS measurements to assess their impact on the occurrence of ghost points. Scanning distance refers to the distance between the sensor and the target; incident angle is the orientation of the target relative to the sensor, *i.e.*, the angle between the laser beam and the normal to the target at the point of incidence of the laser beam; and scanning resolution is defined by the angle between neighboring laser beams, both in the vertical and horizontal plane. The angle in the full (1:1) resolution is 0.009° (in both planes), in the half (1:2) resolution is 0.018° and in the quarter (1:4) resolution is 0.036°. For the FARO® LS 880HE, overlapping laser beams are expected, generating data redundancy at some resolution levels and range [24]. However, this is not meaningful for the data processing explained further. For the TLS measurements, three setups installed on a rotating platform, that mimic canopy-like structures, were used (Figure 1): (i) non-woody materials (hereafter called the leaves setup) are reproduced by squares made of Canson® paper with identical size (*i.e.*, 5 cm × 5 cm) and in a range of different reflectance values (Table 1). This type of paper has a vellum texture minimizing specular reflection; (ii) woody materials (hereafter called the branches setup) are represented by plastic pipes in a range of different diameters and covered by Canson® paper with identical reflectivity; and (iii) overlapping layers of leaves (hereafter called the layers setup) are represented by groups of leaves having identical size and reflectivity (Table 1). Figure 1 shows the three setups on the rotating platform. Distance between objects is fixed to 3 cm in the leaves setup (Figure 1a). In the layers setup (Figure 1b), there are 10 cm between the front and the middle layer and 12 cm between the middle and the back layer. In this setup, the distance between the objects is kept to 3 cm; except for the pairs of objects located on the same layer and close to the center of the platform, where the distance is 5 cm. The distance between the objects in the branches setup (Figure 1c) is 15 cm. The reflectance properties of the materials were measured at the wavelength of the TLS laser beam (*i.e.*, 785 nm), with a FieldSpec® 3 (Analytical Spectral Devices Inc., ASD, Boulder, CO, USA).

Table 1. Overview of the different setups. Reference configuration is underlined.

Variable	Leaves Setup	Layers Setup Values	Branches Setup
Scanning distance	2.5 m, 5 m, 7.5 m, <u>10 m</u> , 12.5 m, 15 m	2.5 m, 5 m, 7.5 m, <u>10 m</u> , 12.5 m, 15 m	2.5 m, 5 m, 7.5 m, <u>10 m</u> , 12.5 m, 15 m
Incident angle	<u>0°</u> , 15°, 30°, 45°, 55°	<u>0°</u> , 15°, 30°, 45°, 55°	<u>0°</u> , 15°, 30°, 45°, 55°
Scanning resolution	1:1, <u>1:2</u> , 1:4	1:1, <u>1:2</u> , 1:4	1:1, <u>1:2</u> , 1:4
Reflectance 785 nm	10%, 50%, 90%	<u>50%</u>	<u>30%</u>
Dimension	5.0 cm × 5.0 cm	5.0 cm × 5.0 cm	5.0 cm, 8.0 cm, 10.0 cm diameter

Additionally, the impact of reflectance and diameter was studied in the leaves and the branches setup, respectively. These values were not changed and were constant during the measurements. A reference configuration was chosen and other variables were changed one at a time. Table 1 provides a technical overview of the different setups.

Figure 1. Measurement setups with canopy-like structures installed on the rotating platform. (a) The leaves setup showing the dimensions of the setup and six objects with reflectance values (left to right): 10%, 50%, 90%, 10% 50% 90%; (b) The layers setup showing 15 objects with reflectance value of 50%; (c) the branches setup showing three objects with a reflectance value of 30%.



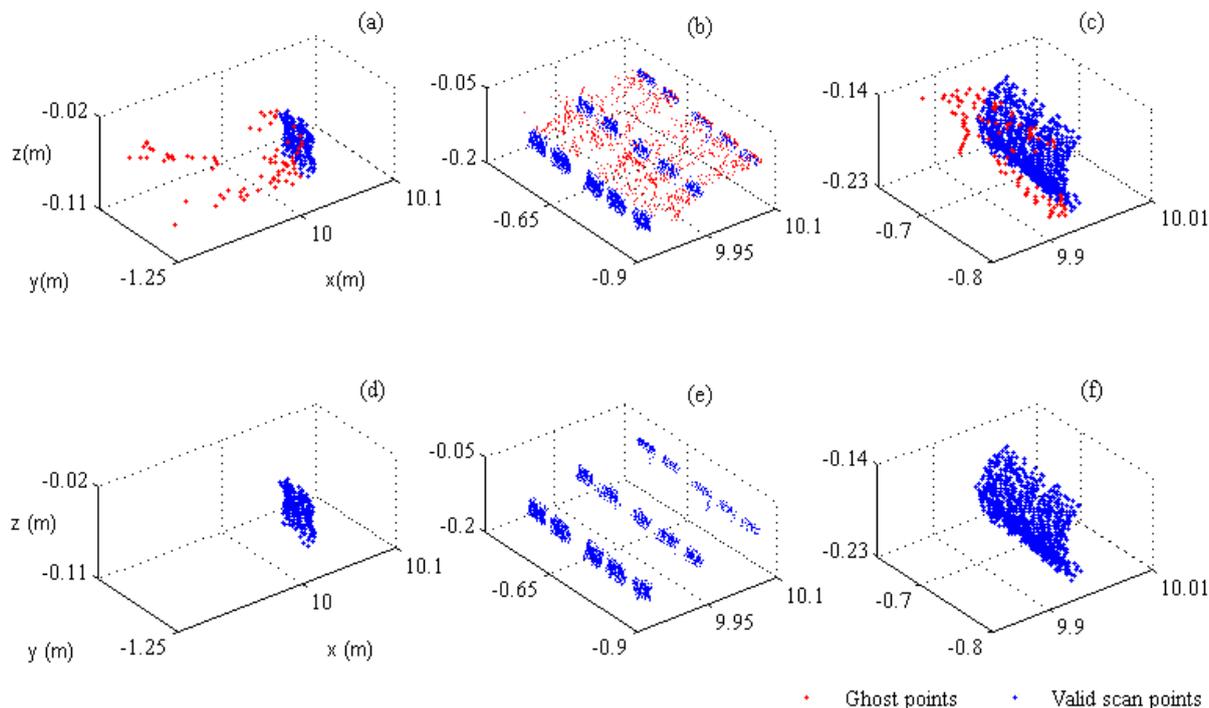
2.3. Manual Extraction of Ghost Points and Ghost Points Ratio Computation

Each point cloud obtained from the corresponding scan was visualized and managed in a 3D environment. Here, ghost points were visually detected in two different ways, depending on the setup. First, in the leaves and the branches setups, the evaluated space within the point cloud comprised the space occupied by each individual object and its immediate surroundings, *i.e.*, after visualization of edges of objects, points further away from these edges, and located more than five times the angle between neighboring beams, defined by the scanning resolution, were not included. This was determined in order to avoid counting irrelevant scan points and/or scan points belonging to adjacent objects. Within this space, and since the objects had known shape and dimension, valid scan points and ghost points were identified (Figure 2a,c). Second, the space examined in the layers setup comprised the larger space occupied by all layers of leaves, *i.e.*, the entire cluster of leaves and immediate surroundings were included (Figure 2b). The scan points belonging to the background were excluded from the analysis. Then, ghost points were manually removed from the considered space (Figure 2d–f) and a ghost points ratio was calculated as:

$$gpr = (sp - vsp)/sp \quad (1)$$

where: *gpr* is the ghost points ratio within the space, with values from zero to one; *sp* the total number of scan points within the space under analysis; and *vsp* the number of valid scan points corresponding to the object. If *gpr* = 0, means no ghost points within the space, while if *gpr* = 1 means there are only ghost points within the space.

Figure 2. Visualization of the point clouds in the space under analysis for (a) the leaves setup; (b) the layers setup; and (c) the branches setup. Valid scan points are shown in blue, ghost points are shown in red; (d–f) show the valid scan points after manual extraction of ghost points, for the leaves setup, the layers setup, and the branches setup, respectively.



With a black background placed as shown in the Figure 1, there will be no measurement errors behind the objects, nor in the middle of objects. This applies to leaves and branches setup, as seen in Figure 2a,c, respectively. As concerns the layers setup this is also true, at least for the noise points originating from the back layer. However, it is rather difficult to determine the origin of noise points that are between layers (Figure 2b).

2.4. Statistical Analysis

Linear regression models were fitted for analyzing the influence of the variables behind the measurement setups (Table 1) on the occurrence of ghost points. The response variable *gpr* (ghost points ratio) was modeled as a function of the variables (*i.e.*, predictor variables) described in Section 2.2. Several models were fitted by the least square method, and these models arise from the combination of the variables that determine each measurement setup. For instance, for the leaves and the branches setup, four predictor variables were involved, thus fifteen models from all possible combinations of these variables were fitted as follows: one model having all four predictor variables; four models having a subset of three different predictor variables; six models having a subset of two different predictor variables and four models having one predictor variable. In the layers setup, three predictor variables were studied, thus seven models were fitted as follows: one model having all three predictor variables; three models having a subset of two different predictor variables and three models having one predictor variable. All fitted models were compared by assessing: (i) the fulfillment of statistical assumptions of linear regression models (*i.e.*, normal distribution and homoscedasticity of

the residuals); (ii) the statistical significance of the estimated parameters; and (iii) the residual standard error. For comparing models having a different number of parameters, e.g., a model having all predictor variables (“full model”) *versus* a model having only a subset of the predictor variables (“reduced model”), a test of hypothesis was carried out based on the *F*-statistics computed using the estimated variance residuals of both models and their respective degrees of freedom. The full models for the leaves, the layers and the branches setups are expressed in Equations (2)–(4), respectively.

$$gprL_i = \beta_0 + \beta_1 dsL_i + \beta_2 rfL_i + \beta_3 rsL_i + \beta_4 aL_i + e_i \quad (2)$$

$$gprLA_i = \beta_0 + \beta_1 dsLA_i + \beta_2 aLA_i + \beta_3 rsLA_i + e_i \quad (3)$$

$$gprB_i = \beta_0 + \beta_1 dsB_i + \beta_2 diB_i + \beta_3 rsB_i + \beta_4 aB_i + e_i \quad (4)$$

where: $gprL_i$, $gprLA_i$ and $gprB_i$ are the ghost points ratio in the i th trial for the object L , LA and B in the leaves, the layers and the branches setups, respectively; ds_i is the distance to the object; rf_i is the reflectance of the object; rs_i is the scanner resolution; a_i is the incident angle; di_i is the diameter of the object; e_i is the i th random error with a Gaussian distribution with mean equal to zero and variance equal to $\sigma_{e_i}^2$; and β_0 , β_1 , β_2 , β_3 and β_4 are parameters of the model.

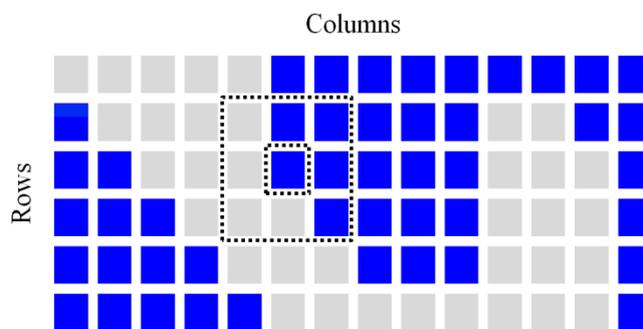
2.5. Extraction of Ghost Points

A distance-based filter was created to detect and remove ghost points. Each point is examined and given a quality value in accordance with a filtering criterion, as explained hereafter. First, in order to properly process the data, the point clouds were projected to a 2D format. According to [25], the most efficient file structure for extracting data is the projection used by [26] for hemispherical photographs. This format, also known as “plate carrée”, is often used for cartographic and GIS data processing. Next, under this new format, the filter was able to compare each scan point with the scan points in an adjacent area. The adjacent area was oriented towards the rows and columns, as they can be seen in the 2D projection of the data (Figure 3). The “kernel size” was the size (expressed in pixels) of the adjacent area used for comparison. For each scan point the filter considers the scan points in the kernel and determines the amount of points that are at approximately the same distance from the scanner as the scan point being evaluated. A scan point was recognized as valid scan point if the difference in distance was smaller than a “distance threshold”. Additionally, there is an “allocation threshold”: the percentage of scan points in the kernel that falls within the distance threshold. The evaluated scan point remains in the point cloud if both the distance and allocation criteria are met. Otherwise, if this quality value was outside the threshold, the scan point was recognized as ghost point and removed. The filter was applied to every single scan point within the point cloud. The standard values assigned to the filter were: kernel size = 3×3 , distance threshold = 2 cm, and allocation threshold = 50%. These values were selected based on the default values used in filters of 3D data processing software (e.g., FARO® Scene).

The configuration of the filter could be modified, by using several combinations of distance and allocation thresholds to detect and remove the ghost points. Improvements to the filtering will be possible by knowing the influence of the different variables of the *gpr*, and after the statistical analysis described in Section 2.4. Finally, *gpr* calculations using both standard and modified configurations of

the filter were compared with the reference *gpr* obtained after the manual extraction of ghost points from Section 2.3.

Figure 3. Schematic of the 2D projection of the data. Blue pixels represent foreground objects and grey pixels may represent the background object or empty space. A single scan point being evaluated and its corresponding adjacent area (kernel) of 3×3 are enclosed within dotted lines.

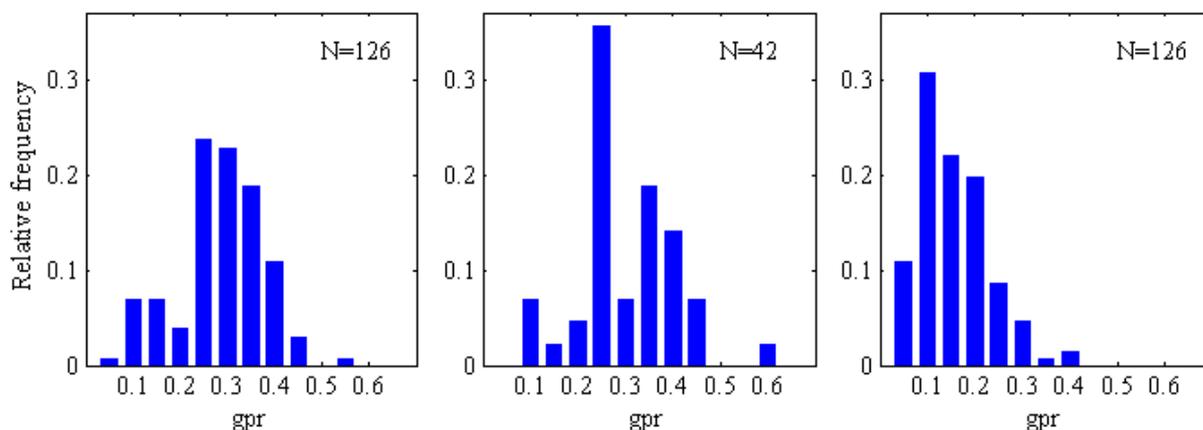


3. Results

3.1. Ghost Points Ratio (*gpr*) Distribution

The ghost points ratio (*gpr*), our response variable, shows different distribution and ranges by each setup, as presented in Figure 4.

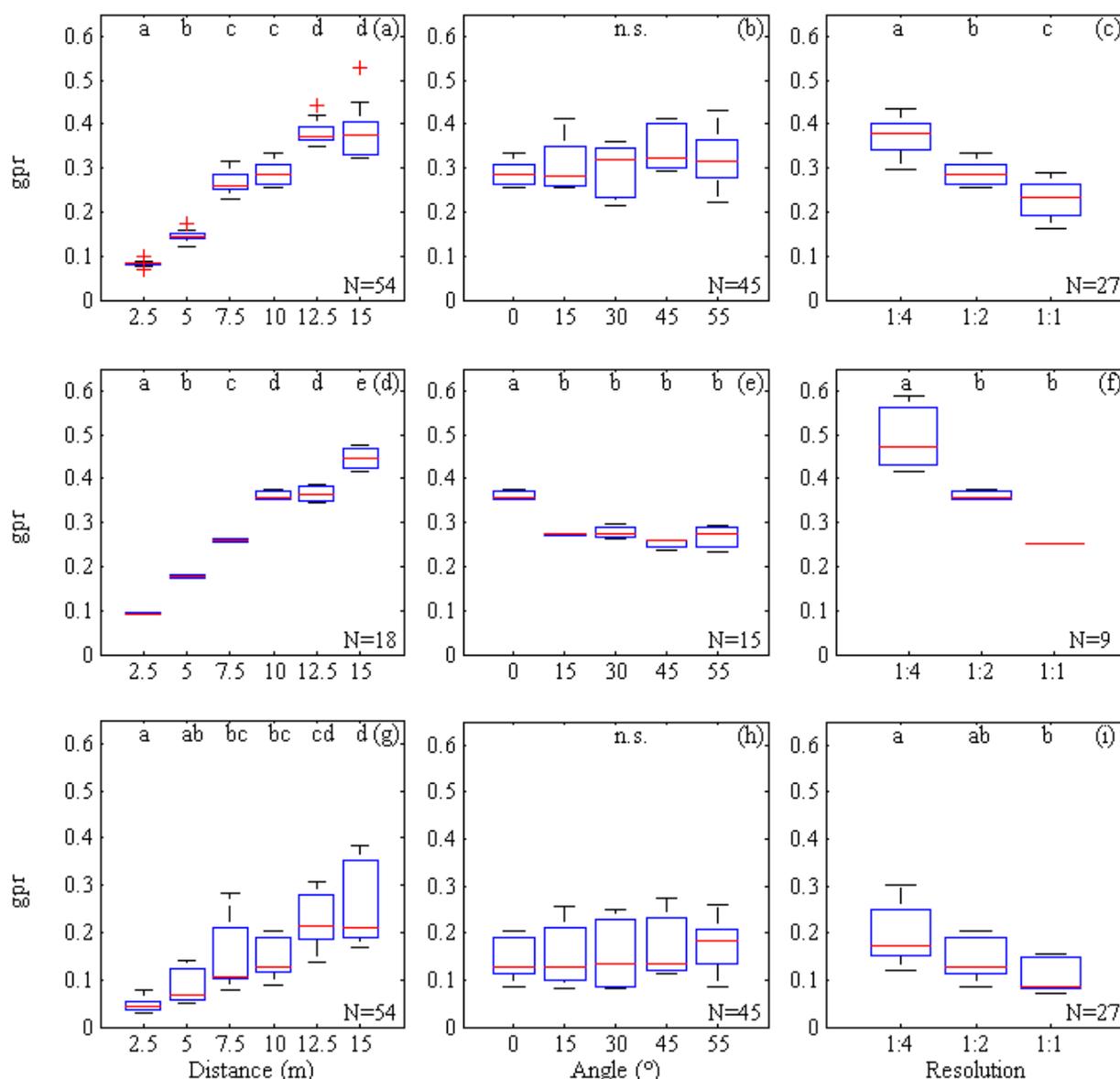
Figure 4. Histograms of ghost points ratio (*gpr*) values over the total number of measurements (*N*). (a) The leaves setup ($N = 126$); (b) the layers setup ($N = 42$) and (c) the branches setup ($N = 126$).



The typical *gpr* in the leaves setup is between 0.25 and 0.3, the smallest *gpr* is below 0.1 and the largest is between 0.5 and 0.55. However, *gpr* above 0.45 is atypical. In general, *gpr* values in the leaves setup are larger than in the branches setup, where typical *gpr* is between 0.1 and 0.15, and *gpr* between 0.05 and 0.25 are frequent. The smallest *gpr* is below 0.05 and the largest is between 0.35 and 0.4. In the layers setup, typical *gpr* is between 0.2 and 0.3, and *gpr* between 0.2 and 0.4 are quite frequent. Similarly to single leaves from the leaves setup, the smallest *gpr* is below 0.1 and the largest

is between 0.5 and 0.6. Figure 5 presents the *gpr* distribution and variability per setup and per variable common to the three measurement setups.

Figure 5. Boxplots displaying the distribution of ghost points ratio (*gpr*) by distance, angle and resolution. The box represents the interquartile range of the data (the 25th and 75th percentiles), and the whiskers represent the inner 10th and 90th percentiles. (a–c) the leaves setup; (d–f) the layers setup; (g–i) the branches setup. Box plots with different letters indicate significant differences between them (Tukey test, $p < 0.05$), and n.s. indicates differences that are not significant ($p > 0.05$). Outliers are represented by plus (+) sign.



In general, the inter-quartile range (IQR) increases from the layers setup to a rather large IQR in the branches setup.

The first predictor variable, *i.e.*, the distance to the object, has a clear influence on the *gpr*, showing an ascending trend in all three measurement setups (Figure 5a,d,g). Median *gpr* values are smaller at shorter distances, becoming greater at longer distances. Variability increases slightly from shorter to longer distances, as well as the extreme values. Although the leaves setup and the layers setup have

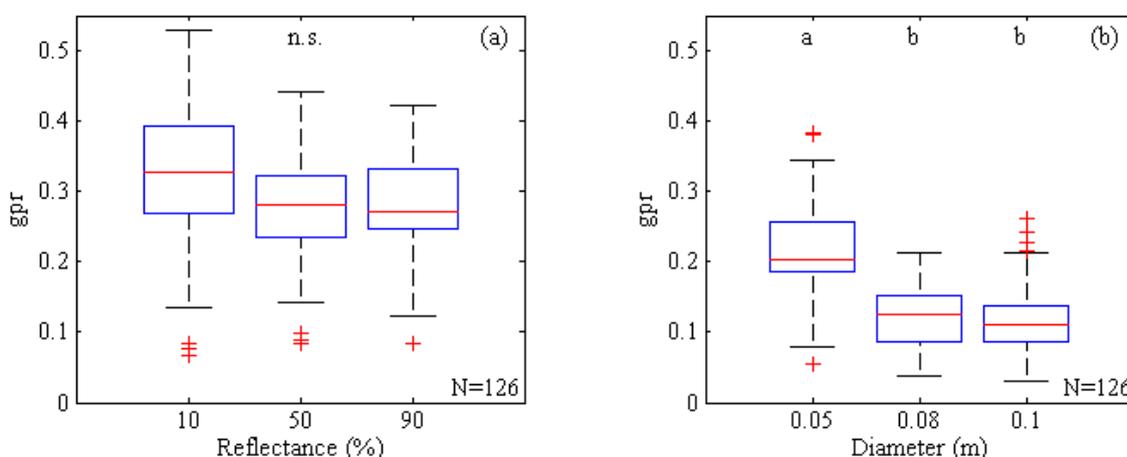
larger *gpr* values than the branches setup (see also Figure 4), a larger variability of *gpr* is observable in the branches setup, for the whole range of different distances.

Secondly; Figure 5b,e,h present the *gpr* calculated per incident angle, showing moderately variable median values across the range of angles evaluated in the experiments. A clear drop in *gpr* in Figure 5e, when the incident angle goes from 0° to 15°, is explained by the organization of objects (*i.e.*, artificial leaves) forming the layers setup. At 0° the objects from the front layer are arranged in line with the objects on the middle layer, from the instrument perspective (Figure 1). Thus, objects on the front layer cover the biggest part of objects on the middle layer. Likewise, the back layer is affected by occlusion from objects on the middle layer. At 15° parts of the objects that were occluded at 0° are now reachable by the laser beam and this is valid for larger incident angles. The ranges of variability show no particular trend, with extreme values slightly farther from the median at incident angle of 55°, and much closer to the median at 0°. A high variability of *gpr* in the branches setup is observable for the whole range of different angles.

Thirdly; Figure 5c,f,i; display the influence of the scanning resolution on the *gpr* values, with a declining trend in *gpr* when increasing resolution. Variability decreases and extreme values are closer to the median in high resolution datasets from the branches and the layers setups. However this is not the case in the leaves setup.

Finally, the effects of leaf reflectance and branch diameter on the *gpr* were also studied (Figure 6).

Figure 6. Boxplots of ghost points ratio *gpr*. (a) Three reflectances; and (b) three diameters in the leaves ($N = 126$) and the branches ($N = 126$) setup, respectively. The box represents the interquartile range of the data (the 25th and 75th percentiles), and the whiskers represent the inner 10th and 90th percentiles. Box plots with different letters indicate significant differences between them (Tukey test, $p < 0.05$), and n.s. indicates differences that are not significant ($p > 0.05$). Outliers are represented by plus (+) signs.



The three boxes of Figure 6a have similar median, with a minor decrease in *gpr* when increasing the reflectance. Variability and extreme values are relatively equal in leaves with medium (50%) and high (90%) reflectance. In turn, variability of *gpr* in less reflective objects (10%) is larger, with extreme values farther from the median. In the branches setup (Figure 6b), the diameter of the objects seems to have a moderate influence on the *gpr*. Median values decrease from small to large diameter. Variability of *gpr* decreases in larger branches. On the medium and large branches (0.08 m and 0.1 m), extreme

values are at similar distance from the median. The smallest branch (0.05 m) presents a larger spread of the data.

3.2. Statistical Analysis

The best reduced models for the leaves, the layers and the branches setup are expressed in Equations (5)–(7), respectively; and are found after fitting the corresponding number of models; *i.e.*, fifteen models for the leaves and the branches setups and seven models for the layers setup.

$$gprL_i = \beta_0 + \beta_1 dsL_i + \beta_2 rfl_i + \beta_3 rsL_i \tag{5}$$

$$gprLA_i = \beta_0 + \beta_1 dsLA_i + \beta_2 rsLA_i \tag{6}$$

$$gprB_i = \beta_0 + \beta_1 dsB_i + \beta_2 diB_i + \beta_3 rsB_i \tag{7}$$

The model for the leaves setup does not include the incident angle (*a*). The relative root mean squared errors (RMSE) of the full model (Equation (2)) and the best reduced model (Equation (5)) are 4.7% and 4.8%, respectively. The *p* value of 0.540 indicates that *H*₀ (variance of the error are equal between the two models) cannot be rejected.

For the layers setup, the relative RMSE of full model (Equation (3)) and the best reduced model (Equation (6)) are 4.5% and 5.6%, respectively. The small *p* value of 3.587e-05 indicates that *H*₀ (variance of the error are equal between the two models) is rejected.

Finally, the model for the branches setup does not include the incident angle (*a*). The relative RMSE of the full model (Equation (4)) and the best reduced model (Equation (7)) are 3.7% and 3.7%, respectively. The *p* value of 0.162 indicates that *H*₀ (variance of the error are equal between the two models) cannot be rejected. Table 2 presents the regression coefficients of the reduced models (*i.e.*, Equations (5)–(7)).

Table 2. Regression coefficients of the best reduced models for the leaves, layers and branches setup.

Model	Variable	Coefficient	Estimate	Standard Error
Leaves (Equation (5))	Intercept	β_0	0.151	0.022
	<i>dsL</i>	β_1	0.026	0.002
	<i>rfl</i>	β_2	−0.051	0.013
	<i>rsL</i>	β_3	−0.173	0.029
Layers (Equation (6))	Intercept	β_0	0.181	0.034
	<i>dsLA</i>	β_1	0.029	0.002
	<i>rsLA</i>	β_2	−0.263	0.047
Branches (Equation (7))	Intercept	β_0	0.219	0.020
	<i>dsB</i>	β_1	0.016	0.001
	<i>diB</i>	β_2	−2.033	0.162
	<i>rsB</i>	β_3	−0.121	0.022

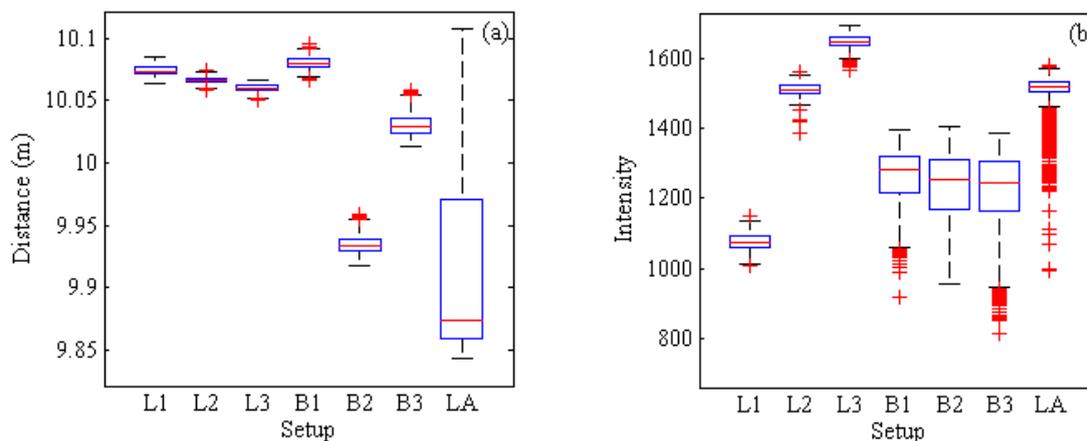
3.3. Extraction of Ghost Points from Point Clouds

A required step, before applying filters for extraction of ghost points, is the examination of the main variables that these filters use to process the data. For this reason, a subdivision of the leaves and the

branches setup is done. The datasets from leaves with 10%, 50% and 90% reflectance were named L1, L2 and L3, respectively. The datasets from branches with 0.05 m, 0.08 m and 0.1 m diameter were named B1, B2 and B3, respectively. Subdivision of the layers setup was not possible, thus, this dataset was named LA.

Figure 7 plots the variability in distance and intensity of valid scan points. Variability in distance of valid scan points from individual leaves within the leaves setup is below 10 mm with extreme values close to the median. The variability within the branches setup presents an expected increase in the IQR when increasing the diameter, with similar distance to extreme values from the median. The layers setup presents a particularly increased IQR, which was also expected, given the variables distance to the different layers of leaves. In this case, the behavior of the variability is assumed to be equivalent to the variability of L2 (Figure 7a). In conclusion, the recorded distances between sensor and objects presents reduced variability, meaning that the distance-based filter is applicable to these datasets.

Figure 7. Boxplot of (a) distance and (b) intensity of valid scan points per setup. L1: leaf reflectance 10%, L2: leaf reflectance 50%, L3: leaf reflectance 90%, B1: branch diameter 0.05 m, B2: branch diameter 0.08 m, B3: branch diameter 0.1 m, LA: layers setup. The box represents the interquartile range of the data (the 25th and 75th percentiles), and the whiskers represent the inner 10th and 90th percentiles. Reference configuration from Table 1 was used: scanning distance = 10 m; incident angle = 0°; scanning resolution = 1:2. Outliers are represented by plus (+) signs.



Contrasting behavior is perceived for intensity, where it can be seen from Figure 7b that the variability of the intensity of valid scan points in the leaves setup is high, with extreme values far from the median and various outliers. An even greater variability is seen in the branches setup. The layers setup presents a similar situation as the leaves setup with a greater number of outliers. In summary, attempting to define appropriate thresholds on an intensity-based filter becomes problematic, even under the controlled conditions of this experimental study. However, if the aim is to develop an effective filter using this property, intensities can be calibrated with external references.

The mean percentage of ghost points recognized by the distance-based filter with the standard configuration, is presented in Table 3. The *gpr* from manual extraction of ghost points is used as reference data.

Table 3. Mean percentage and standard deviation (SD) of ghost points detected using a distance-based filter. The standard configuration was used for the filter: kernel size = 3×3 , distance threshold = 2 cm, and allocation threshold = 50%.

Setup	Mean (%)	SD
L1	71.52	10.2
L2	66.37	13.1
L3	67.00	10.4
LA	67.54	13.2
B1	62.26	6.4
B2	65.88	12.4
B3	63.67	7.8

It is observable that the distance-based filter underestimates the amount of ghost points in all setups, ranging from 71.5% of detection in the less reflective leaf to 62.3% in the smallest branch, meaning that, in general, the filter is not able to detect all ghost points that were identified manually.

Given the results from the statistical analysis in Section 3.2, evidencing that the distance between sensor and object is the main variable influencing ghost points occurrence, the filter was modified to efficiently detect and remove the ghost points from point clouds collected at different distances. Thus, the configuration (*i.e.*, distance and allocation threshold) of the filter was changed, considering the average points per configuration on each experimental setup and optimizing the percentage rate of detection. Table 4 gives a detailed overview of the values used to optimize the distance-based filter and the improvements on the ghost points detection, contrasting the amount of ghost points identified after manual extraction with the ghost points recognized by the filter.

Table 4. Overview of the distance-based configuration at different distances and percentage of ghost points detected using the optimized distance-based filter. d (m): distance threshold, a (%): allocation threshold. L1: leaf reflectance 10%, L2: leaf reflectance 50%, L3: leaf reflectance 90%, B1: branch diameter 0.05 m, B2: branch diameter 0.08 m, B3: branch diameter 0.1 m, LA: layers setup. SD: standard deviation.

Setup		2.5 m	5 m	7.5 m	10 m	12.5 m	15 m	Mean (%)	SD
L1	d (m)	0.01	0.01	0.01	0.015	0.015	0.015	97.7	8.6
	a (%)	50	50	62.5	62.5	62.5	75		
L2	d (m)	0.006	0.007	0.007	0.012	0.018	0.02	102.3	10.1
	a (%)	50	50	62.5	62.5	75	75		
L3	d (m)	0.005	0.006	0.007	0.008	0.01	0.012	100.5	3.7
	a (%)	50	50	62.5	62.5	62.5	62.5		
LA	d (m)	0.006	0.007	0.008	0.009	0.01	0.017	101.7	2.5
	a (%)	50	50	50	50	50	50		
B1	d (m)	0.008	0.008	0.008	0.015	0.015	0.015	101.0	8.8
	a (%)	62.5	62.5	62.5	62.5	75	75		
B2	d (m)	0.009	0.01	0.01	0.018	0.02	0.02	99.5	4.6
	a (%)	62.5	62.5	62.5	62.5	75	75		
B3	d (m)	0.009	0.01	0.01	0.018	0.02	0.02	98.3	7.8
	a (%)	62.5	62.5	62.5	62.5	75	75		

In spite of the slight under and over estimation, the results after using the optimized distance-based filter confirm significant improvements in ghost points detection for all setups, with standard deviation below 10%. The largest overestimation is presented in L2 with 2.3 percentage points. The largest underestimation is presented in L1, with -2.3 percentage points.

4. Discussion

Ghost points were manually identified from the point clouds to determine the reference *gpr* and enable filter performance evaluation in data quality improvement. While this manual method presented some disadvantages, such as, subjective selection of ghost points and increased processing time, it was the most direct way to generate a reference dataset. Results show that *gpr* is mainly influenced by: (i) the distance between the sensor and objects being evaluated; (ii) scanning resolution and (iii) the angle of incidence. The latter variable, however, has no significant influence on the leaves and the branches setup. These results are in agreement with the findings by [12] on their modeling of edge loss from one solid object. Main outcomes after manual extraction of ghost points confirmed that *gpr* was relative to the size of objects. This could be seen on the branches setup, where objects with three different diameters were tested. With higher total area, the relative amount of edge points becomes smaller and thus the number of ghost points decreases. This may be a problem with thin structures in forest canopies (*i.e.*, twigs), in cases where the object has less angular width than the width of the laser beam, it is possible that no measurement reports the true position of the object, but instead all scan points may be classified as ghost points [22]. The fact that variability of *gpr* in the branches setup is higher than in the leaves and in the layers setups (Figure 5) is also explained by the different size of branches used in this study.

The impact of object reflectance on ghost points is verified, as in [18,20,21]. The effect on *gpr* if scanning resolution increases is comparable to the effect on *gpr* if objects are larger. Similar results presented in [12], suggest that attention must be paid to the beam size and divergence, in order to avoid redundancy on the data caused by overlapping when measuring in high resolution at reduced distances. Conversely, if the beam divergence is larger than the resolution, information loss can occur and an accurate representation of the measured object becomes difficult [27]. On the present study, the latter situation is less likely to occur. Yet, overlapping is present on the following scenarios: from the beam exit to ± 8 m when using scanning resolution 0.036° ; to ± 47 m using 0.018° ; and at all ranges with 0.009° . The *gpr* is not affected by this redundancy because is indicating a simple ratio between valid and ghost points detected on an object.

It is worth noting that, even though the selection of the space to be analyzed was restricted to the immediate surroundings of the object, the interaction of the laser beam with the background might have had an effect on the amount of ghost points outside the defined space. Continuous-wave operating laser scanners, like the one used in this experiment, have the inherent problem of producing ghost points anywhere along the line of sight of the laser beam [10]. Furthermore, the distance between the object and the background is also important. According to [12], when the background surface is less reflective or further away from the object, less energy is reflected from the background surface, and the ghost points range is shifted toward the front surface, as can be seen in Figure 2. In that sense, reducing the reflectivity of the back surface will make the second response too weak to influence the

positioning result, as can occur at canopy level, in areas where there is no background. On the other hand, findings in the behavior of ghost points occurrence using time-of-flight laser scanners indicate that this type of instruments will always work in the first response mode, disregarding the signal strength between the object and background signals [22,28,29].

Multiple regression analysis was useful in these experimental situations where predictor variables were controlled. Several predictor variables were investigated simultaneously, finding that more than one key predictor variable influenced the response (*i.e.*, *gpr*) in each setup. The analyses for the leaves and the branches setup indicate that corresponding reduced models are statistically equal. Both reduced models do not include incident angle as predictor variable, meaning that the effect of incident angle in the *gpr* is not significant. In contrast, after the analysis in the layers setup, we infer that the difference in *gpr* estimation between the full and reduced models is significant, thus, all predictor variables in the full model influence *gpr*. Incident angle is significant in this setup, most probably as a consequence of the symmetric organization of objects conforming the layers, as mentioned in the description of Figure 5e.

Although the use of TLS intensity data as reflectance information has been proved to be effective to separate foliage from wood in trees and other applications [8,30], it is not useful to define a fixed set of filter parameters under the conditions of the present study. Intensities should be calibrated with an external reference if the aim is to develop an effective intensity-based filter. Radiometric calibration would be needed to find a sequence of corrections that convert the raw intensity information into a value proportional or equivalent to target reflectance [31]. Furthermore, and according to the manufacturer's information, the digital value of the intensity also depends on the parameters of the analogue to digital conversion. Nevertheless, the radiometric calibration does not necessarily guarantee an effective detection of ghost points if they are within the same range of intensity values as the valid scan points of the object under observation.

A distance-based filter is suitable and applicable to improve the quality of the datasets of this study. Reducing the distance threshold lead us to decrease the allocation threshold as well, in order to identify the ghost points at different distances. This becomes useful for this type of indoor experiments with solid objects. However, if the filtering routine is applied to real forest canopy data, probably a different criteria has to be chosen. Indeed, increasing the distance threshold with increasing distance between object and sensor, and at the same time increasing the allocation threshold (Table 4), give more probabilities that adjacent points from actual objects are detected by the filter. Given the complexity of a forested scene, where all types of objects, distances and inclinations are present at once; a preferred approach to apply these results should consider: (i) selection of the appropriate scanning resolution; (ii) *in situ* determination of species composition, assessment of leaf reflectance and leaf angle distribution; (iii) a first classification of the point cloud data in photosynthetic and non-photosynthetic material, based on the intensity values and preferably including the aforementioned radiometric calibration. After that, the distance-based filter can be applied to the photosynthetic material, taking into consideration the reflectance properties and with average thresholds similar to those presented in Table 4. It is recommended to consider a customized filtering in function of the zenith angle when including incident angle for ghost points detection (*i.e.*, layers setup). The outcome can then be used for vegetation structure analysis, when the objective is to accurately model, for example, the branches and trunks within forest canopies [7]. In summary, considering the aforementioned recommendations, the developed

method in this study will not only remove ghost points at the edge of canopy perimeters when gaps are distinguishable [23], but will also delete ghost points from scenes with overlapping objects and consequently improve continuous-wave TLS data quality.

5. Conclusions

This study gives evidence that the ghost points' occurrence in point cloud data, collected using a continuous-wave TLS, is affected by a number of variables depending on object properties and their organization in the space. Improvements to the quality of TLS data from different experimental setups, including single and overlapping objects, was achieved through elimination of these ghost points using filtering techniques. An intensity-based filter is not suitable for this experiment without considering radiometric calibration. Applying distance-based filter algorithms, using the appropriate combination of thresholds is recommended to detect and remove ghost points. There is certainly an inevitable search for a trade-off while configuring filtering algorithms, in order to avoid deletion of valid scan points and, at the same time, improve ghost points recognition. Hence, even though it would be enough to remove ghost points up to a level if post-processing algorithms work reliably, a rather intensive filtering was preferred for the processing of ghost points. This comes with the risk that valid scan points may be deleted, but additional processing procedures such as voxelization, may assist to correct the misclassified spaces that might have been created by using such intensive filtering. This research was a first step in the deletion of ghost points from simulated canopy scenes and its transferability to natural environments is not straightforward. Further research is necessary, in order to validate the use of the optimized filter on complex scenes, such as real forest canopies, where all types of objects, distances and inclinations are present at once. Scanning technology, *in situ* measurements within the 3D scene and classification after radiometric calibration are factors to bear in mind in order to obtain reliable results. This will deliver an enhanced data quality before further processing, allowing to build a more accurate representation of the 3D structure of forest canopies.

Acknowledgments

The authors would like to thank Belspo STEREO II Programme (Hyperforest Project SR/00/134) for funding this study. We express our gratitude to Dirección de Investigación y Desarrollo, Universidad de La Frontera, Chile, DIUFRO Project No. DI11-0018 for supporting Christian Salas contributions. We would also like to thank two anonymous reviewers for their constructive comments and suggestions.

Author Contributions

The contribution of the authors to this work is as stated: Renato Cifuentes^{1,2,3,4,5,6,7,8}, Dimitry Van der Zande^{1,3,5,6,7}, Christian Salas^{3,4,5,7}, Jamshid Farifteh^{1,6,7,8} and Pol Coppin^{1,7,8}. ¹ Experimental design; ² data collection; ³ data processing; ⁴ development of algorithm; ⁵ statistical analysis and interpretation; ⁶ manuscript preparation and writing; ⁷ paper review; ⁸ project coordination.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Parker, G.G. Structure and Microclimate of Forest Canopies. In *Forest Canopies*; Lowman, M.D., Nadkarni, N.M., Eds.; Academic Press: San Diego, CA, USA, 1995; pp. 73–106.
2. Harper, J.L. Canopies as populations. In *Plant Canopies: Their Growth, Form and Function*; Russell, G., Marshall, B., Jarvis, P.G., Eds.; Cambridge University Press: Cambridge, UK, 1989; pp. 105–128.
3. Niklas, K.J. Evolutionary walks through a land plant morphospace. *J. Exp. Bot.* **1999**, *50*, 39–52.
4. Porter, J.R. Modules, models and meristems in plant architecture. In *Plant Canopies: Their Growth, Form and Function*; Russell, G., Marshall, B., Jarvis, P.G., Eds.; Cambridge University Press: Cambridge, UK, 1989; pp. 143–159.
5. Seidel, D.; Fleck, S.; Leuschner, C.; Hammett, T. Review of ground-based methods to measure the distribution of biomass in forest canopies. *Ann. For. Sci.* **2011**, *68*, 225–244.
6. Zheng, G.; Moskal, L.M. Computational-Geometry-Based Retrieval of Effective Leaf Area Index Using Terrestrial Laser Scanning. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3958–3969.
7. Newnham, G.; Armston, J.; Muir, J.; Goodwin, N.; Tindall, D.; Culvenor, D.; Püschel, P.; Nyström, M.; Johansen, K. Evaluation of Terrestrial Laser Scanners for Measuring Vegetation Structure. In *CSIRO Sustainable Agriculture Flagship*; Manuscript ID: EP124571; CSIRO Research Publications Repository, 2012. Available online: <https://publications.csiro.au/rpr/pub?pid=csiro:EP124571> (accessed on 20 May 2014).
8. Côté, J.-F.; Widlowski, J.-L.; Fournier, R.A.; Verstraete, M.M. The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial lidar. *Remote Sens. Environ.* **2009**, *113*, 1067–1081.
9. Balduzzi, M.A.F.; van der Zande, D.; Stuckens, J.; Verstraeten, W.W.; Coppin, P. The Properties of Terrestrial Laser System Intensity for Measuring Leaf Geometries: A Case Study with Conference Pear Trees (*Pyrus communis*). *Sensors* **2011**, *11*, 1657–1681.
10. Hebert, M.; Krotkov, E. 3D measurements from imaging laser radars: How good are they? *Image Vis. Comput.* **1992**, *10*, 170–178.
11. Mills, J.; Barber, D. Geomatics Techniques for Structural Surveying. *J. Surv. Eng.* **2004**, *130*, 56–64.
12. Tang, P.; Akinci, B.; Huber, D. Quantification of edge loss of laser scanned data at spatial discontinuities. *Autom. Constr.* **2009**, *18*, 1070–1083.
13. Magney, T.S.; Eusden, S.A.; Eitel, J.U.H.; Logan, B.A.; Jiang, J.; Vierling, L.A. Assessing leaf photoprotective mechanisms using terrestrial LiDAR: Towards mapping canopy photosynthetic performance in three dimensions. *New Phytol.* **2014**, *201*, 344–356.
14. Van der Zande, D.; Stuckens, J.; Verstraeten, W.W.; Mereu, S.; Muys, B.; Coppin, P. 3D modeling of light interception in heterogeneous forest canopies using ground-based LiDAR data. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 792–800.
15. Lichti, D.; Jamtsho, S. Angular resolution of terrestrial laser scanners. *Photogramm. Rec.* **2006**, *21*, 141–160.

16. Kwak, D.-A.; Lee, W.-K.; Kafatos, M.; Son, Y.; Cho, H.-K.; Lee, S.-H. Estimation of effective plant area index for South Korean forests using LiDAR system. *Sci. China Life Sci.* **2010**, *53*, 898–908.
17. Soudarissanane, S.; Lindenbergh, R.; Menenti, M.; Teunissen, P. Scanning geometry: Influencing factor on the quality of terrestrial laser scanning points. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 389–399.
18. Cang, Y.; Borenstein, J. Characterization of a 2-D Laser Scanner for Mobile Robot Obstacle Negotiation. In Proceedings of the 2002 IEEE International Conference on Robotics and Automation, Washington, DC, USA, 11–15 May 2002; Volume 3, pp. 2512–2518.
19. Boehler, W.; Vicent, M.B.; Marbs, A. Investigating laser scanner accuracy. In Proceedings of the XIXth CIPA Symposium, Antalya, Turkey, 30 September–4 October 2003.
20. Tang, P.; Huber, D.; Akinci, B. A Comparative Analysis of Depth-Discontinuity and Mixed-Pixel Detection Algorithms. In Proceedings of the Sixth International Conference 3-D Digit, Imaging Model (3DIM 2007), 21–23 August 2007; pp. 29–38.
21. Adams, M.D.; Probert, P.J. The Interpretation of Phase and Intensity Data from AMCW Light Detection Sensors for Reliable Ranging. *Int. J. Robot. Res.* **1996**, *15*, 441–458.
22. Tuley, J.; Vandapel, N.; Hebert, M. Analysis and Removal of Artifacts in 3-D LADAR Data. In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 2203–2210.
23. Vaccari, S.; van Leeuwen, M.; Calders, K.; Coops, C.; Herold, M. Bias in lidar-based canopy gap fraction estimates. *Remote Sens. Lett.* **2013**, *4*, 391–399.
24. Cifuentes, R.; van der Zande, D.; Farifteh, J.; Salas, C.; Coppin, P. Effects of voxel size and sampling setup on the estimation of forest canopy gap fraction from terrestrial laser scanning data. *Agric. For. Meteorol.* **2014**, *194*, 230–240.
25. Jupp, D.L.B.; Culvenor, D.S.; Lovell, J.L.; Newnham, G.J.; Strahler, A.H.; Woodcock, C.E. Estimating forest LAI profiles and structural parameters using a ground-based laser called “Echidna”. *Tree Physiol.* **2009**, *29*, 171–181.
26. Andrieu, B.; Sohbi, Y.; Ivanov, N. A Direct Method to Measure Bidirectional Gap Fraction in Vegetation Canopies. *Remote Sens. Environ.* **1994**, *50*, 61–66.
27. Pesci, A.; Teza, G.; Bonali, E. Terrestrial Laser Scanner Resolution: Numerical Simulations and Experiments on Spatial Sampling Optimization. *Remote Sens.* **2011**, *3*, 167–184.
28. Lichti, D.D.; Gordon, S.J.; Tipdecho, T. Error models and propagation in directly georeferenced Terrestrial Laser Scanner Networks. *J. Surv. Eng.* **2005**, *131*, 135–142.
29. Matthies, L.; Bergh, C.; Castano, A.; Macedo, J.; Manduchi, R. Obstacle Detection in Foliage with Ladar and Radar. In *Robotics Research, STAR 15*; Dario, P., Chatila, R., Eds.; Springer-Verlag Berlin: Heidelberg, Germany, 2005; pp. 291–300.
30. Rosell Polo, J.R.; Sanz, R.; Llorens, J.; Arnó, J.; Escolà, A.; Ribes-Dasi, M.; Masip, J.; Camp, F.; Gràcia, F.; Solanelles, F.; *et al.* A tractor-mounted scanning LIDAR for the non-destructive measurement of vegetative volume and surface area of tree-row plantations: A comparison with conventional destructive measurements. *Biosyst. Eng.* **2009**, *102*, 128–134.

31. Kaasalainen, S.; Krooks, A.; Kukko, A.; Kaartinen, H. Radiometric Calibration of Terrestrial Laser Scanners with External Reference Targets. *Remote Sens.* **2009**, *1*, 144–158.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).