Improved 3D Stem Mapping Method and Elliptic Hypothesis-Based DBH Estimation from Terrestrial Laser Scanning Data

WenFang Ye 1, Chuang Qian 1,*, Jian Tang 1, Hui Liu 1, XiaoYun Fan 1, Xinlian Liang 2 and HongJuan Zhang 3

1 GNSS Research Center, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; wenfangYe@whu.edu.cn (W.Y.); tangjian@whu.edu.cn (J.T.); Loweliu@whu.edu.cn (H.L.);
2017286180031@whu.edu.cn (X.F.)
2 Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, 02431 Masala, Finland; xinlian.liang@fgi.fi
3 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; hongjuanzhang@whu.edu.cn
* Correspondence: qc_gnss@whu.edu.cn; Tel.: +86-132-3550-6666

Received: 20 December 2019; Accepted: 19 January 2020; Published: 21 January 2020

Abstract: The detailed structure information under the forest canopy is important for forestry surveying. As a high-precision environmental sensing and measurement method, terrestrial laser scanning (TLS) is widely used in high-precision forestry surveying. In TLS-based forestry surveys, stem-mapping, which is focused on detecting and extracting trunks, is one of the core data processing tasks and the basis for the subsequent calculation of tree attributes; one of the most basic attributes is the diameter at breast height (DBH). This article explores and improves the methods for stem mapping and DBH estimation from TLS data. Firstly, an improved 3D stem mapping algorithm considering the growth direction in random sample consistency (RANSAC) cylinder fitting is proposed to extract and fit the individual tree point cloud section. It constructs the hierarchical optimum cylinder of the trunk and introduces the growth direction into the establishment of the backbone buffer in the next layer. Experimental results show that it can effectively remove most of the branches and reduce the interference of the branches to the discrimination of trunks and improve the integrity of stem extraction by about 36%. Secondly, a robust least squares ellipse fitting method based on the elliptic hypothesis is proposed for DBH estimation. Experimental results show that the DBH estimation accuracy of the proposed estimation method is improved compared with other methods. The mean root mean squared error (RMSE) of the proposed estimation method is 1.14 cm, compared with other methods with a mean RMSE of 1.70, 2.03, and 2.14 cm. The mean relative accuracy of the proposed estimation method is 95.2%, compared with other methods with a mean relative accuracy of 92.9%, 91.9%, and 90.9%.

Keywords: terrestrial laser scanning (TLS); 3D stem-mapping; diameter at breast height (DBH); robust least square elliptic fitting

1. Introduction

Assessing the spatial organization of trees within the forest is a key objective for both forest managers and researchers [1]. Forest inventory measures the structural parameters on a sample plot, which provides an important basis for obtaining the quantity, quality, stand structure, and growth pattern of the wood, and determines the biomass, stem volume, forest carbon cycle, biodiversity, and changes in these attributes [2,3]. The individual tree is the basic measurement unit for forest inventory, and the stem is an important component of it. Stem volume is the most targeted variable
in commercial forest management [4]. In addition, stems are closely related to the extraction of forest attributes, including tree locations [5,6], heights [7,8], diameters at breast height (DBH) [9–11], stem curves [12], crown widths [13,14], wood volumes [15,16], biomass [17–19], and so on. In the conventional forest resources surveys, the detailed stem attributes are measured by investigators by using caliper and altimeter in the field [2,7]. This survey method, which relies on manual measurement, is time-consuming and can only obtain limited data.

With the development of remote sensing, light detection and ranging (LiDAR) becomes one of the most promising remote sensing technologies for estimating various biophysical properties and structure parameters of forests [20,21]. Terrestrial laser scanning (TLS) is an effective tool for obtaining detailed structural information under the tree crown at the regional and forest holding level, because it can obtain high accuracy and high-density point clouds under the forest canopy [2]. Based on TLS, the geometry attributes can be measured efficiently, which are based on stem mapping that detects and extracts the stem from the point cloud. Many researches have studied and proposed some solutions and algorithms for stem detection and extraction from the TLS data in forest areas. These methods can be divided into three main categories: (1) Image-based methods; (2) 2D-based methods; and (3) 3D-point-based methods.

The first method groups the pixels in the range image based on local properties, such as the distance or surface curvature, and then uses the image process approaches to extract laser points belonging to the tree stems [22–24].

The second method divides the point cloud into slices in the vertical direction, and then conducts shape detection on the slices in order to detect the stem laser points. Tree stems are identified by point clustering and the detecting shape algorithm, e.g., circle fitting [10], cylinder fitting [25], and the Hough-transform-based circle fitting method [26]. This method mainly considers the two-dimensional geometric characteristic, which is simple and of lower computational cost. However, stem point detection becomes a problem when there are branches overlapping in the layer. In addition, a digital terrain model (DTM) needs to be produced before constructing the slice at a certain height, because the knowledge of the terrain is necessary.

The last method extracts and classifies 3D point clouds directly based on geometric, physical, and other features, which can be further divided into point-based methods and segment-based methods. Point-based stem extraction methods, which detect the stem points based on the points’ features, have been studied by many researchers. Liang et al. [27] proposed a method to identify stem points based on flatness and normal direction. Lalonde et al. [28] introduced a method that classifies the points according to three geometric features, including scatter, linear, and surface with the Bayesian classification method. The linear points are treated as the stem points to be extracted. However, the geometric features that vary with the radius of the neighborhood need to be calculated at each point by its surrounding points [29], which will result in a heavy computational burden. In addition to the points’ geometric features, some scholars have also proposed the point-based stem extraction method based on other point characteristics, such as the intensity and radiometric [30], to extract the stem points in point clouds. However, the intensity of laser reflection is affected by many factors, and the calibration of intensity is complicated [31]. Segment-based methods perform a classification based on point clouds’ segments. They consider the geometric features of the 3D point clouds’ segments instead of the point, which can reduce the calculation and improve the calculation speed [32]. These methods divide point clouds into fragments by segmentation algorithms, such as shape detection [33] and region growing [34], and then classify them according to the characteristics and shapes of fragments. Burt et al. [35] proposed a cylindrical fitting method based on random sample consistency (RANSAC) to extract the stem up to the position of first branching. Liang et al. [36] proposed a method that constructed a series of overlapping cylinders along the stem to compose a full trunk model. Based on the concepts of tree topology and cover sets, Raumonen et al. [37] proposed a method to divide point clouds into stems and branches by using surface patches and checking the local connectivity of a moving surface region. Most of these methods assume the tree stem as segmented cylindrical models and extract the stem points by the fitting method, such as direct cylinder fitting and random sample consistency (RANSAC). However, in the process of cylindrical
fitting, the points of branches will affect the fitting result and even result in the failure of fitting, which will lead to some parts of the stem not being able to be detected. Especially in the upper half of the tree stem, where the stem is thinner, the number of stem points is fewer. The points of the branches will seriously affect the results of cylindrical fitting. Therefore, the existing methods can extract the lower part of the stem, but it is difficult to extract the stem near the top of the tree, which affects the integrity of the trunk extraction.

After the tree stem is extracted, the attribute parameters of the tree need to be further estimated. DBH is one of the most important factors in forest surveying that provides a basis for calculating the stem volume and constructing a growth model. Thus, the accurate estimation of DBH is important for high precision forest surveying. The main task is to estimate the optimal DBH parameters from the point cloud of the trunk at the corresponding height. Many articles have proposed many DBH estimation methods, such as linearized or nonlinear least square circle fitting [3,7,24], Hough-transform [26], cylinder fitting [15,38], random sample consensus (RANSAC) algorithm [8,39], and random Hough transform [31,40]. Most of these methods model the stem profiles as a circle and fit the diameter parameter from the stem points at the breast height. Although, the shape of the stem profiles is very similar to a circle, it is not standard, which introduces errors into the fitting process. Moreover, although DBH is called a diameter parameter, it is determined by the perimeter of the tree stem at the breast height, which is measured in the forest survey (the DBH is estimated by dividing the perimeter by $\pi$). Therefore, we do not have to use a circular model to get the DBH parameter and are able to choose a more accurate model to fit and get a more accurate perimeter, which can estimate a more accurate DBH parameter. The extracted stem points may include a few outliers and noise points, which will also degrade the fitting accuracy. Hence, we should reduce the effect of these branch points in fitting.

To summarize, to improve the integrity of individual stem extractions for stem mapping and the accuracy of DBH estimation, we improved an RANSAC-based stem extraction method and proposed an improved DBH estimation method. The main contribution included the following: (1) The sectional RANSAC cylindrical fitting method is improved by considering the growth direction. The growth direction of the stem is calculated and utilized to exclude the branches points in the establishment of the stem buff, which improves the integrity of stem extraction. (2) A robust least squares ellipse algorithm is proposed to fit the stem section and reduce the influence of branches’ points in fitting, which can improve the accuracy of the DBH estimation.

The rest of this paper is organized as follows. The automatic extraction process of tree parameters, including the trunk extraction method and robust least squares ellipse algorithm, is introduced in Section 2. Field experiment and experimental results are presented in Section 3. The discussion and conclusions are drawn in Section 4.

2. Materials and Methods

2.1. Research Plot and Data Description

Three research plots’ data were studied, two of which were measured in the Hubei Province, China, and the other one was from the open data in the TLS benchmarking project.

Wuhan is located in the east of the Hubei Province, China (30.54°N, 114.31°E). Gongan is located in the south-central part of the Hubei Province, China (30.07°N, 112.32°E), which is a typical plain lake area with a 32% forest coverage. The TLS datasets used in this study were obtained from a Populus euramerican forest plot in Yangjiachang Town, Gongan, 40 by 100 m, and a metasequoia forest plot in Ziyang Park, Wuhan city, 30 by 50 m. Two study areas located in Hubei Province are shown in Figure 1. The open data in the TLS benchmarking project are distributed in a southern boreal forest in Evo, Finland (61.19°N, 25.11°E). Each plot has a fixed size, 32 by 32 m. The main tree species are Scots pine (*Pinus sylvestris* L.), Norway spruce (*Picea abies* L. Karst.), and silver (*Betula pendula* Roth) and downy (*Betula pubescens* Ehrh.) birches. In order to distinguish, the three data are named plot Gongan, plot Wuhan, and plot Finland.
In order to obtain the best visibility for all the trees, multi-scan mode was used to collect forest point cloud data. In the multi-scan mode, several scans are performed within the forest plots to collect more detailed point clouds to represent the sample plot. Two scans were performed in the Wuhan plot and three scans in the Gongan plot. These scans were co-registered by using spherical targets that were manually placed throughout the plot.

The point cloud data were obtained by the laser scanner FARO Focus3D X 130 (FARO Technologies, Inc., USA) in 2019. The instrument specifications of the scanner are shown in Table 1. In Table 2, the detailed data descriptions of 3 plots are listed, which includes the mean DBH, sum of trees, stand conditions, density, and main species. All the trees in three plots were used for a stem point extraction experiment. Parts of trees with a measured DBH in plots Wuhan and Gongan and all trees in plot Finland were used for the DBH extraction experiment. The trees involved in DBH extraction experiments were measured by caliper in the field and marked sequentially in Figure 2. In the DBH extraction experiment, the experimental samples in Gongan were 46 trees planted along the trails in the woodland, and the experimental samples in Wuhan were 23 randomly selected trees.

### Table 1. Technical specifications of the Focus3D X130.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data acquisition speed</td>
<td>976,000 points/s</td>
</tr>
<tr>
<td>Maximum range</td>
<td>153 m</td>
</tr>
<tr>
<td>Maximum resolution</td>
<td>Vertical: 0.009°</td>
</tr>
<tr>
<td></td>
<td>Horizontal: 0.009°</td>
</tr>
<tr>
<td>Field of view angle</td>
<td>300 × 360°</td>
</tr>
<tr>
<td>Scanner line speed</td>
<td>97 Hz</td>
</tr>
<tr>
<td>Measurement accuracy</td>
<td>± 2 mm</td>
</tr>
<tr>
<td>Wavelength</td>
<td>1,550 nm</td>
</tr>
</tbody>
</table>

### Table 2. Plot descriptions of the 3 sample plots.

<table>
<thead>
<tr>
<th>Plot</th>
<th>Mean DBH (cm)</th>
<th>Sum of trees</th>
<th>Stand conditions</th>
<th>Density (stems/ha)</th>
<th>Main Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gongan</td>
<td>26.91</td>
<td>160</td>
<td>leaf-off</td>
<td>400</td>
<td>Poplar</td>
</tr>
<tr>
<td>Wuhan</td>
<td>33.49</td>
<td>122</td>
<td>leaf-on</td>
<td>813</td>
<td>Metasequoia</td>
</tr>
<tr>
<td>Finland</td>
<td>16.66</td>
<td>84</td>
<td>leaf-on</td>
<td>820</td>
<td>Scots pine</td>
</tr>
</tbody>
</table>
2.2. TLS Data Preprocessing

2.2.1. Co-Registration, Ground Filtering, and TLS Data Thinning
Point cloud data preprocessing includes co-registration, ground filtering, and TLS data thinning. Co-registration is used to convert the different scans into the same coordinate system by using the common points. In multi-station scanning, there are five visible spherical targets between each station and the coordinates of the reference targets are used as the common points in the point cloud co-registration of different scans. Then, we completed the point cloud co-registration with SCENE, which is a software packaged with the laser scanner FARO. According to the processing report of SCENE, the co-registration error was less than 6 mm.

Ground filtering is a necessary preprocessing process for isolating the individual trees. The scanned forest point clouds not only include trees but also a large number of ground points, which will hinder the detection and extraction of tree point clouds. In addition, ground filtering is a prerequisite for disconnecting the stem from the ground, and non-ground point clouds are easier to achieve in individual tree segmentation. The cloth simulation filter method is a ground filtering method based on a physical process. It utilizes the nature of cloth and modifies the physical process of cloth simulation to adapt to point cloud filtering [41]. Zhang et al. [31] used the cloth simulation filter method to filter the ground. The ground points and non-ground points were also segmented by the cloth simulation filter method in our study. This function is integrated into the open source software Cloud Compare, which can complete the ground filtering operation with the software.

Finally, the non-ground point cloud was sampled in the pre-processing. Due to data redundancy, the processing takes a long time and data thinning is a necessary process to improve processing efficiency. We streamlined the data by the voxel grid filter, which down-sampled the original point cloud data and each voxel was 1 cm$^3$ in size, which could sample the point cloud without losing accuracy in tree detection [42].

2.2.2. Individual Tree Extraction

The individual tree is the basic measurement unit. Individual tree clouds are the basis of automatic extraction of tree attributes. The next step is the segmentation of the non-ground cloud into individual tree points. Many individual tree segmentation algorithms have been proposed, such as connected component labeling [40], mean shift method [43], point cloud directly segmentation [44], and Euclidean distance clustering [45]. Euclidean distance clustering and connected component labeling were used for individual tree extraction in this research. Euclidean distance clustering in a Euclidean sense can be implemented by making use of a 3D grid subdivision of the space by a k-dimensional tree (kd-tree) structure in order to find the nearest neighbors. It is suitable for the segmentation of the sample plot with low tree density. The label connected components algorithm can be used to find the connected components within organized point cloud data. The input point cloud is divided into small parts by a 3D gird, which is deduced from the octree structure. Based on the point cloud part, connected component labeling can be used to connect point cloud parts and complete the segmentation of the tree stem from stratified point clouds [40]. The detailed procedures can be found in [40,45]. Finally, the isolated points and point clusters were then detected and deleted in the segmentation process.

2.3. Stem Point Extraction Based on the Growth Direction

After individual trees are extracted, the individual tree not only contains the trunk but also the canopy, branches, leaves, and other components. For 3D stem mapping, the 3D stem points should be extracted further from the individual tree points. Some methods treat the trunk as a vertical cylinder with pole-like characteristics and identify it by using a single vertical cylinder model [46]. However, a single cylinder model cannot describe the trunk exactly, which will degrade the stem extraction result. Therefore, improved methods are proposed by utilizing the multi-cylinders model instead of the single cylinder model, which can describe the trunk better [12].

Based on the multi-cylinder model, the tree stem points can be extracted according to the fitting results. The RANSAC algorithm has the characteristics of fewer iterations and a strong ability of resisting gross errors [39], which is a robust fitting method and was utilized. In the fitting process, a stem buffer is created firstly, which will include other tree components, such as branches, that would
introduce the errors in fitting. If a stem buffer is created using the growth direction constraint, which can remove the interference components (e.g., branches) in the point cloud of the candidate area in advance, the accuracy of the RANSAC calculation results can be enhanced.

Figure 3 shows the process for stem extraction. Firstly, we divided the tree point cloud into sections based on height and sorted them from the bottom to the top. The first stem buffer was established containing the points of bottom section. The RANSAC algorithm was applied for fitting optimal cylindrical parameters that describe the buffer best. Then, the central axis of the cylinder was calculated by the cylindrical parameters. Considering that the growth direction of the adjacent sections of the tree stem is similar, the next stem buffer was established with the central axis parameter of the former section, which can reject part of the points of other components from the stem buffer and improve the RANSAC fitting result. Considering that the growth direction of trees is not fixed, the frustum-shaped buffer zone was more robust than the cylindrical buffer zone. It can reduce the misjudgment of the stem caused by the deviation of the growth direction and allow more main stems to be included in the buffer to be calculated. This process was repeated until the last section.

The stem buffer establishment with the central axis parameter is the key to stem point extraction based on the growth direction. As shown in Figure 4(a), the cylindrical axis parameter, \( L(l, m, n) \), can be defined and calculated by the two points on the cylindrical axis, which are \( P(x, y, z) \) and \( P_0(x_0, y_0, z_0) \) in the figure. \( R \) is the radius of the cylinder. The steps of the next stem section extraction considering the \( L \) and \( R \) are as follows:

1. As Figure 4(b) shows, a frustum buffer was established by the coordinates of a point on the axis, \( L \) and \( R \). The frustum buffer can be regarded as the result of the angle of rotation of the fitted RANSAC cylinder around the central axis of the cylinder;
2. As shown in Figure 4(d), the pink cylinder is represented by the RANSAC cylinder that best fits the trunk of this section. The inside point of the RANSAC cylindrical model is considered the stem point;
3. As shown in Figure 4(e), the yellow area in the figure is a frustum buffer created based on the cylinder parameters provided by the RANSAC cylinder; and
4. Figure 4(f) shows that the central trunk components of tree clustering were extracted layer by layer with fixed steps. By splicing the extraction results of each layer together, the splicing results can be regarded as the trunk of an individual tree.

![Figure 3. Flow chart of the segment-based method for detecting tree stem points.](image-url)
Figure 4. Sketch map of the stem point extraction. (a) Schematic diagram of the cylinder and its parameters; (b) frustum schematic; (c) individual tree; (d) random sample consistency (RANSAC) cylinder fitting for the first section of tree points; (e) frustum-shaped buffer established based on cylindrical parameters; (f) stem points extracted.

2.4. Robust Least Squares Elliptic Fitting for DBH Estimation

After stem extraction, the DBH can be estimated from the stem points at breast height. There have been many DBH methods proposed, such as linear least square (Landau algorithm) circle fitting [3,10], nonlinear least squares (Gauss Newton) circle fitting [7], crescent moon method proposed by Kiraly and Brolly [47], RANSAC circle detection [39], Hough transform, and random Hough transform [26,31,40]. However, most of them are based on the assumption that the stem section is circular.

Although the shape of stem profiles is similar to a circle, there are still small differences. Compared with the circular model, the elliptic model can describe the shape of stem profiles, which also includes the circular model, and hence can improve the fitting parameters’ accuracy of stem profiles. Moreover, the elliptic model cannot directly calculate the diameter. It needs to calculate the perimeter of the ellipse according to the model parameters and then divide it by pi to obtain the DBH parameters, which is also the same for the actual measurement method of the current forestry survey.

The accuracy of the fitting is reduced because the extracted stems cannot completely remove the branch points. At the same time, the outliers of the measurement will also affect the fitting results. The DBH estimation method should reduce the effect of these errors on the estimation results. Robust estimation is an estimation method widely used in the measurement field, which can alleviate the influence of outliers [48]. IGG III is one of the widely used robust estimation methods.
This paper applies robust least squares based on the IGG III method to ellipse fitting. The IGG III weight function can effectively remove the interference of the coarse data points on the result by assigning zero weights to the coarse domain data and enhance the robustness of the algorithm.

For sliced DBH point cloud data, we regarded them as ellipses and used robust least squares elliptic fitting to find the best fitting for the given set of points. The mathematical representation of use is the conics equation of the ellipse, which is:

\[ \text{Ellipse} = Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, \quad (1) \]

In order to avoid the trivial solution, \([ABCEDE]^T = 0_{6 \times 1}\), the constraint \(F = -1\) is applied. Note that \(F\) does not depend on edge point \((x, y)\), so Equation (2) is a least squares problem instead of a total least squares problem [48]:

\[ f(a, x) = a^T x - 1 = Ax^2 + Bxy + Cy^2 + Dx + Ey - 1 = 0, \quad (2) \]

where ellipse parameters are given by \(a = [ABCEDE]^T\), and the data point is \(X = [x^2 \ y^2 \ xy]^T\). The fitting of a general conic may be approached by minimizing the sum of squared algebraic distances:

\[ g(a) = \sum_{i=1}^{n} f^2(a, x_i). \quad (3) \]

According to Equation (3), a least squares fit is performed on the quantities that do not satisfy the equation; then, the equations can be written in the matrix form (Equation (4)):

\[ V = GX - L, \quad (4) \]

where,

\[
V = \begin{bmatrix}
v_1 \\
v_2 \\
\vdots \\
v_N \\
\end{bmatrix},
G = \begin{bmatrix}
x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 \\
x_2^2 & x_2y_2 & y_2^2 & x_2 & y_2 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_N^2 & x_Ny_N & y_N^2 & x_N & y_N \\
\end{bmatrix},
X = \begin{bmatrix}
A \\
B \\
C \\
D \\
E \\
\end{bmatrix},
L = \begin{bmatrix}
1 \\
1 \\
1 \\
1 \\
1 \\
\end{bmatrix}.
\]

According to the principle of least squares, there are the following formulas:

\[ X = (G^T \hat{P} G)^{-1}(G^T \hat{P} L), \quad (6) \]

where \(\hat{P}\) is the equivalent weight matrix of the edge points of an ellipse. In order to improve the accuracy of the ellipse parameter solution, we introduced the robust method and selected the robust weight function as follows:

\[ w_i = \begin{cases} 
\frac{1}{k_0(k_1-|\hat{v}_i|)} & |\hat{v}_i| \leq k_0 \\
\frac{|\hat{v}_i|}{|\hat{v}_i|(k_1-k_0)} & k_0 \leq |\hat{v}_i| \leq k_1, \\
0 & |\hat{v}_i| \geq k_1 
\end{cases} \; (7) \]

where \(v_i\) is the residuals’ error of the measurement, \(\sigma_0\) is the standard deviation of the measurement error, \(|\hat{v}_i| = \frac{|v_i|}{\sigma_0}\) is the standardized residuals, and \(k_0\) and \(k_1\) are the modulation coefficients of the robust threshold. According to Equation (7), the normal data is retained, the weight of suspicious data is reduced, and bad data is eliminated. It can effectively eliminate the effect of gross errors on ellipse fitting.

According to elliptic coefficients, it can be calculated as follows: Ellipse center coordinate \((x_0, y_0)\), two semi-axes \(a_e\) and \(b_e\), and orientation \(\theta\). The perimeter of the ellipse and the DBH can be calculated by the following formula:

\[ C_{\text{ellipse}} = 4(a_e + b_e) - 4\pi + \frac{0.1218(a_e - b_e)^2}{(a_e + b_e)^2 + 2.0a_e b_e} \left(\frac{a_e b_e}{a_e + b_e}\right), \quad (8) \]

\[ \text{DBH}_{\text{ellipse}} = C_{\text{ellipse}}/\pi. \quad (9) \]
The proposed robust least square ellipse fitting (RLSEF) method and the other three commonly used DBH estimation methods, random sampling consistency circle fitting (RANSAC_CF), random sampling consistency ellipse fitting (RANSAC_EF), and random Hough transform (RHT), were compared. In the RLSEF and RANSAC_EF methods, the perimeter of the ellipse is calculated first by the model parameters, and DBH is calculated by the perimeter divided by \( \pi \). In the RANSAC_CF and RHT methods, the DBH can be calculated directly by the model parameters.

### 2.5. Evaluation Methods

In tree extraction, two types of error were quantified in this study. Type I error, or omission, is represented by the number of trees not extracted. Type II error, or commission, is the number of trees falsely extracted [36]. In accordance with the TLS benchmarking project [49], the accuracy of tree extraction was assessed by the completeness, correctness, and mean accuracy of detection. The completeness is defined by the percentage of the extracted trees in the field. The correctness measures the percentage of the trees extracted correctly. The mean accuracy is the joint probability that an extracted tree randomly chosen is correct. They were calculated as:

\[
\text{Completeness} = \frac{n_{\text{match}}}{n_{\text{ref}}},
\]

(10)

\[
\text{Correctness} = \frac{n_{\text{match}}}{n_{\text{extr}}},
\]

(11)

\[
\text{Mean accuracy of detection} = \frac{2n_{\text{match}}}{(n_{\text{ref}} + n_{\text{extr}})}.
\]

(12)

Ravaglia et al. used the maximum extraction height that is defined as the height above the ground of the highest diameter of the stem to evaluate three tree stem diameter algorithms [50]. In our study, the integrity of stem extraction was calculated further to evaluate the ability of algorithms for stem extraction by the maximum extraction height, \( H_{\max} \), and tree height, \( H_{\text{tree}} \), in the scanned points cloud. The \( H_{\max} \) is calculated by \( z_{\max} \) and \( z_{\text{ground}} \), where \( z_{\max} \) is the height of the extracted highest off-ground stem point and \( z_{\text{ground}} \) is the height at which the tree intersects the ground. \( H_{\text{tree}} \) is calculated by the \( z_{\text{tree}} \) and \( z_{\text{ground}} \), and \( z_{\text{tree}} \) is obtained manually from the individual tree point cloud:

\[
H_{\max} = z_{\max} - z_{\text{ground}}, H_{\text{tree}} = z_{\text{tree}} - z_{\text{ground}}.
\]

(13)

The integrity is calculated as:

\[
\text{Integrity} = \frac{H_{\max}}{H_{\text{tree}}}.\]

(14)

Mean DBH is the diameter corresponding to the average basal area of the dominant tree species, which is a basic index reflecting the forest roughness. The accuracy of DBH estimation has also been addressed by other authors. The accuracy of the estimated DBH at tree level was evaluated by calculating \( R^2 \), bias, root mean squared error (RMSE), relative bias, relative RMSE, and relative accuracy in this study. The relative bias, relative RMSE, and relative accuracy were calculated according to the following formula, where \( d_i \) is the estimated DBH, \( D_i \) is the measured DBH, and \( n \) is the total number of trees in forest sample plots:

\[
\text{Bias} = \frac{1}{n} \sum_{i=1}^{n} (d_i - D_i),
\]

(15)

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (d_i - D_i)^2},
\]

(16)

\[
\text{Relative Bias} = \frac{\text{Bias}}{\text{mean DBH}}.
\]

(17)
\[ \text{relative RMSE} = \frac{\text{RMSE}}{\text{mean DBH}} \]  
\[ \text{relative accuracy} = 1 - \text{relative RMSE}. \]

3. Results

3.1. Individual Tree Extraction Result

Firstly, the ground point and the non-ground point were segmented by the cloth simulation filter in the open source software Cloud Compare. For plot Gongan, the Euclidean clustering algorithm in the PCL (Point Cloud Library) was used to segment ground points. The planting distance of each tree in plot Gongan was about four meters, so the search radius of the nearest neighbor search was set as 0.1 m, which can effectively divide the forest into individual trees. By setting the minimum cluster size to 500, outlier point clusters could be deleted.

For plot Finland and plot Wuhan, we used the label connected components algorithm that has been integrated into Cloud Compare to find the connected components within organized point cloud data. When the octree level equals 10, trunks, shrubs, and weeds show better separability and the amount of data increased substantially when the octree level is bigger than 10 [40]. We used a 3D grid to extract the connected components. This grid was deduced from the octree structure, and the octree level was set to 10. As can be seen from the Figure 5, the non-ground point cloud can be divided into individual trees’ point clouds after segmentation.

In order to analyze the results of tree extraction, the number of trees in the original data was counted manually as in the reference. The numbers of trees that were extracted correctly and incorrectly were counted separately. Table 3 lists the accuracy assessments for three plots. The tree density of plot Gongan is 400 stems/ha. All trees were correctly extracted. In plot Wuhan, 11 of the extracted trees belong to type II error. Eight of them contained some clusters from another tree, and three of them contained several trees. In plot Finland, six of the extracted trees belonged to Type II error and all of them contained several trees. The statistical results are shown in Table 3. According to the standard in [49], the individual tree extraction is effective.
Figure 5. Individual tree extraction result. (a) Plot Gongan; (b) plot Wuhan; (c) plot Finland.

Table 3. Accuracy of trunk extraction using the terrestrial laser scanning (TLS) data.

<table>
<thead>
<tr>
<th>Plot</th>
<th>Reference</th>
<th>Mapped</th>
<th>Type I Error</th>
<th>Type II Error</th>
<th>True Culms</th>
<th>Completeness (%)</th>
<th>Correctness (%)</th>
<th>Mean accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gongan</td>
<td>160</td>
<td>160</td>
<td>0</td>
<td>0</td>
<td>160</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Wuhan</td>
<td>122</td>
<td>118</td>
<td>2</td>
<td>11</td>
<td>107</td>
<td>87.70</td>
<td>90.68</td>
<td>89.17</td>
</tr>
<tr>
<td>Finland</td>
<td>84</td>
<td>78</td>
<td>2</td>
<td>6</td>
<td>72</td>
<td>85.71</td>
<td>92.31</td>
<td>88.89</td>
</tr>
</tbody>
</table>

3.2. Stem Point Extraction Result
For the separated individual tree point cloud, we used the improved 3D stem-mapping algorithm by considering the growth direction in RANSAC multi-cylinder fitting to extract the stem point. The results, without considering the growth direction, were compared. The maximum extraction heights for the piecewise RANSAC cylinder fitting method and piecewise RANSAC cylinder fitting method constrained by the growth direction were calculated, respectively. To analyze the integrity of stem mapping, the visible heights of the tree stems were measured manually from the original point cloud.

Extraction of the individual tree and stem were performed on all trees on three plots. Some representative extraction results were selected to reflect our algorithm and the stem extraction results of the two methods for stem extraction, which is shown in Figure 6. The left figure is the piecewise RANSAC cylinder fitting method, and the right figure is our improved algorithm. The black points in the figure are the tree points, and the red points are the extracted stem points. As shown in Figure 6 (a) and (b), the stem extraction results of two typical trees in plot Finland by the two methods were compared to show the improvement. As shown in Figure 6 (c) and (d), the stem extraction results of two randomly selected trees in plots Gongan and Wuhan by the two methods were compared.

From Figure 6, it is obvious that the maximum extraction heights by the improved method are higher in three plots, which means the integrity of the extracted stem points is improved. For quantitative analysis, the mean maximum extraction heights and mean tree heights were calculated by all the trees of three plots, respectively. The mean integrities in three plots were calculated further. As shown in Table 4, the statistic of the mean tree height and the mean maximum extraction height of each plot are listed. In three plots, the integrity of stem extraction is improved by 31%, 19.8%, and 56.8%. The mean integrity of the proposed method in the three plots is 84.0%, compared with the other method with a mean integrity of 62.3%. The mean improvement is 36%.
Figure 6. Comparison of piecewise random sample consistency (RANSAC) cylinder fitting and piecewise RANSAC cylinder fitting constrained by growth direction. (a) Plot Finland; (b) plot Finland; (c) plot Gongan; (d) plot Wuhan. (Left: sectional RANSAC fitting. Right: RANSAC fitting combined with growth direction. Stem points are shown in red and tree points are shown in black).

Table 4. Stem extraction with two algorithms.

<table>
<thead>
<tr>
<th>Plot</th>
<th>Mean tree height (m)</th>
<th>Mean maximum extraction height (m)</th>
<th>Mean integrity (%)</th>
<th>Mean maximum extraction height (m)</th>
<th>Mean integrity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gongan</td>
<td>27.03</td>
<td>17.22</td>
<td>63.71</td>
<td>22.57</td>
<td>83.50</td>
</tr>
<tr>
<td>Wuhan</td>
<td>17.82</td>
<td>11.99</td>
<td>67.28</td>
<td>14.37</td>
<td>80.64</td>
</tr>
<tr>
<td>Finland</td>
<td>13.69</td>
<td>7.67</td>
<td>56.03</td>
<td>12.03</td>
<td>87.87</td>
</tr>
</tbody>
</table>
3.3. DBH Estimation Result

A point cloud with a height of 1.275 to 1.325 m was selected for slicing, called a DBH point cloud, and projected onto the XOY plane. Instead of the circle model, the elliptical model was applied in the robust least square elliptic fitting algorithm for DBH estimation.

To evaluate the accuracy of the DBH estimation, the true DBHs of some trees were measured by the diameter ruler, which are shown in the Figure 2. There are 45 trees in the Gongan, 23 trees in the Wuhan, and 77 trees in the Finnish plot, with true DBHs analyzed to evaluate the accuracy of DBH estimation.

The proposed RLSEF method and the other three commonly used DBH estimation methods, RANSAC_EF, RANSAC_CF, and RHT, were compared. We used the coefficient of determination ($R^2$), bias, RMSE, relative bias, relative RMSE, and relative accuracy to measure the accuracy of DBH estimation.

Table 5. Accuracy comparison of four methods for estimating the diameter at breast height (DBH).

<table>
<thead>
<tr>
<th>Plot</th>
<th>Algorithms</th>
<th>$R^2$</th>
<th>RMSE (cm)</th>
<th>Bias (cm)</th>
<th>RMSE%</th>
<th>Bias%</th>
<th>Relative accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gongan</td>
<td>RLSEF</td>
<td>0.952</td>
<td>1.114</td>
<td>1.188</td>
<td>4.1</td>
<td>4.4</td>
<td>0.959</td>
</tr>
<tr>
<td></td>
<td>RANSAC_EF</td>
<td>0.871</td>
<td>1.2</td>
<td>1.400</td>
<td>4.5</td>
<td>5.2</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td>RANSAC_CF</td>
<td>0.838</td>
<td>2.161</td>
<td>1.683</td>
<td>8</td>
<td>6.3</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>RHT</td>
<td>0.777</td>
<td>1.645</td>
<td>1.920</td>
<td>6.1</td>
<td>7.1</td>
<td>0.939</td>
</tr>
<tr>
<td>Wuhan</td>
<td>RLSEF</td>
<td>0.932</td>
<td>1.185</td>
<td>0.950</td>
<td>3.5</td>
<td>2.8</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td>RANSAC_EF</td>
<td>0.856</td>
<td>2.263</td>
<td>1.903</td>
<td>6.8</td>
<td>5.7</td>
<td>0.932</td>
</tr>
<tr>
<td></td>
<td>RANSAC_CF</td>
<td>0.86</td>
<td>2.377</td>
<td>2.044</td>
<td>7.1</td>
<td>6.1</td>
<td>0.929</td>
</tr>
<tr>
<td></td>
<td>RHT</td>
<td>0.739</td>
<td>2.546</td>
<td>2.052</td>
<td>7.6</td>
<td>6.1</td>
<td>0.924</td>
</tr>
<tr>
<td>Finland</td>
<td>RLSEF</td>
<td>0.978</td>
<td>1.12</td>
<td>0.876</td>
<td>6.7</td>
<td>5.3</td>
<td>0.933</td>
</tr>
<tr>
<td></td>
<td>RANSAC_EF</td>
<td>0.955</td>
<td>1.643</td>
<td>1.235</td>
<td>9.9</td>
<td>7.4</td>
<td>0.901</td>
</tr>
<tr>
<td></td>
<td>RANSAC_CF</td>
<td>0.973</td>
<td>1.54</td>
<td>1.247</td>
<td>9.2</td>
<td>7.5</td>
<td>0.908</td>
</tr>
<tr>
<td></td>
<td>RHT</td>
<td>0.897</td>
<td>2.233</td>
<td>1.951</td>
<td>13.4</td>
<td>11.7</td>
<td>0.866</td>
</tr>
</tbody>
</table>

Figure 7 shows the scatter plot of TLS-estimated DBH and field-measured DBH in three plots. From the graph, the goodness-of-fit of the RLSEF method is the highest. The R-squared of this method is also the highest of the four methods. The results show that the RLSEF method produces lower mean errors for the three plots than with the other three algorithms.
Figure 7. Scatter plot of estimated diameter at breast height (DBH) values.
From Table 5, we can see that the robust least squares ellipse fitting performs the best out of the four algorithms. The mean RMSE of the proposed estimation method is 1.14 cm, compared with the other methods’ mean RMSE values of 1.70, 2.03, and 2.14 cm. The mean relative accuracy of the proposed estimation method is 95.2%; compared with other methods, the mean relative accuracy is 92.9%, 91.9%, and 90.9%. Stem profiles are elliptical, and as the DBH increases, this elliptical shape becomes more pronounced. The average DBH of the trees in plots Wuhan and the Gongan are both greater than 26 cm and the relative RMSE and relative accuracy of the robust least square ellipse fitting algorithm in these two plots is higher than that in Finland. This proves that the elliptical model can describe the stem profiles better. The result is also verified in Figure 8, which shows the fitted results of three plots. Obvious ellipse characteristics are seen from the experimental stem profiles of Wuhan and Public Security. The average DBH of the Finnish plot is less than 17 cm. Furthermore, each index in Table 5 shows that robust least squares ellipse fitting is the best fitting method among them. It should be mentioned that, due to the voting mechanism of the random Hough transform, the detected circle results have a certain false positive rate. For example, the best fitting circles of two trees in plot Gongan are not unique while using the RHT method, and the fitted circle is significantly smaller, which is shown in Figure 9.
4. Discussion

In the stem extraction experiment, the mean integrities of the proposed method in three plots are higher than those with the cylindrical RANSAC fitting method. Especially in plot Finland, the mean integrity of the proposed method is 87.7%; compared with the cylindrical RANSAC fitting method, the mean integrity is 56.03%. The experimental results show that the improved 3D stem-mapping method can improve the maximum extraction height significantly, especially in the Finland plot where the trees have dense branches. This is because when the candidate points for fitting contain many branch points, the cylindrical RANSAC fitting may fail to fit the cylinder, which will lead to the failure of stem extraction. Because the stem points are decreasing while branch points are increasing in the upper part of the trunk, the fitting method cannot work effectively. By introducing the growth direction into the establishment of the candidate points buffer in the next layer, most of the branch points can be removed, which will help the fitting process. Hence, the improved 3D stem mapping method is effective for the stem extraction of trees with dense branches.

In the DBH estimation experiment, the mean RMSE of the proposed estimation method is 1.14 cm; compared with other methods, it is 1.70, 2.03, and 2.14 cm. The mean relative accuracy of the proposed estimation method is 95.2%, and compared with other methods, it is 92.9%, 91.9%, and 90.9%. The experimental results show that the elliptic model fitting method can improve the accuracy of the DBH estimation, which verifies that the DBH calculated based on perimeter is also equivalent to the existing measurement method. Even when the shape of the stem section is close to a circle in the Finland plot, the elliptic model fitting method can still obtain the equivalent DBH estimate accurately compared with the circular model methods. The robust least squares elliptic fitting shows the best fitting effect and robustness, and it can fit all scatter points in the maximum range. The results indicate that extracting DBH through the robust least squares ellipse fitting method is feasible and shows better accuracy. This is because when the circular model cannot describe the stem section well, the fitting model will introduce the errors in DBH estimation. An elliptic model can include the possibility of a circular model and be more robust to fit the stem section. Furthermore, the robust least squares elliptic fitting method can reduce the weight of gross errors and noise points, which can improve the accuracy of fitting compared with the RANSAC method.

Furthermore, the DBH estimation accuracy of random Hough transform is relatively lower than other methods, because rasterization of the 2D layer at DBH will cause a loss of accuracy. The first reason is that random Hough circle detection after pixel transformation decreases the accuracy and data format conversion reduces the availability of data. The second reason is that circle fitting does...
not adapt to the actual shape of the tree breast height, thus resulting in a decline of the detection accuracy.

5. Conclusion

This paper improved a 3D stem mapping algorithm to collect individual tree information from plot-level terrestrial laser scanning data. This method can effectively remove most of the branches and reduce the interference of the branches in the discrimination of trunks. This paper also presented a robust least square ellipse fitting method that better fits the actual shape of the stem section and can improve the accuracy of DBH estimation.

The methods were tested by the data from forest plots in Gongan, Wuhan, and Finland. According to the experimental results, the following conclusions were drawn:

- With the combination of the tree growth direction and the sectional RANSAC cylindrical fitting method, the integrity of stem extraction is improved compared with that without considering the growth direction. Therefore, the growth direction is an important attribute of the stem and should be considered in stem extraction methods;
- Compared with the circular model, the elliptic model can describe the stem section better, and the DBH estimation accuracy based on the elliptic fitting method is higher than that based on circular fitting methods; and
- Compared with other algorithms, the robust least square ellipse fitting method has performed best in the estimation of DBH, with an average RMSE of 1.14 cm, which can greatly improve the accuracy of DBH estimation based on TLS.

At present, the entire process requires the use of multiple software products and is complicated and time consuming. In future research, automation is an important task in order to apply the methods in actual forestry surveys.

**Author Contributions:** Algorithm materials, data analysis and writing—original draft preparation, W.Y.; experimental design and writing—review and editing, Q.C.; experimental data collection, X.F.; review, J.T., H.L., X.L., and H.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was financially supported by the Fundamental Research Funds for the Central Universities (No. 2042019KF0034).

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


© 2019 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 (CC BY) license (http://creativecommons.org/licenses/by/4.0/).