APPROXIMATION OF VOLUME AND BRANCH SIZE DISTRIBUTION OF TREES FROM LASER SCANNER DATA

Pasi Raumonen (1), Sanna Kaasalainen (2), Mikko Kaasalainen (1), and Harri Kaartinen (2)

(1) Department of Mathematics, Tampere University of Technology
P.O. Box 553, FI-33101 Tampere, Finland, pasi.raumonen@tut.fi
(2) Remote Sensing and Photogrammetry, Finnish Geodetic Institute
P.O.Box 15, FI-02431 Masala, Finland

KEY WORDS: LIDAR, point cloud, trees, classification, size approximation, carbon cycle

ABSTRACT:

This paper presents an approach for automatically approximating the above-ground volume and branch size distribution of trees from dense terrestrial laser scanner produced point clouds. The approach is based on the assumption that the point cloud is a sample of a surface in 3D space and the surface is locally like a cylinder. The point cloud is covered with small neighborhoods which conform to the surface. Then the neighborhoods are characterized geometrically and these characterizations are used to classify the points into trunk, branch, and other points. Finally, proper subsets are determined for cylinder fitting using geometric characterizations of the subsets.

1 INTRODUCTION

The total above-ground volume, the branch size distribution, and other size and shape parameters of trees are of economic and scientific interests. For example, carbon cycle estimations of forests require the branch size distribution of trees for the accurate determination of branch decay times (carbon release as a function of time). One way to approximate these parameters is to use dense terrestrial laser scanner produced point clouds (Pfeifer et al., 2004; Maas et al., 2008; Rutzinger et al., 2010). In many of these methods, the point cloud is mapped to a voxel space, where it is segmented into branches by, e.g., mathematical morphology (Gorte and Pfeifer, 2004). Then cylinders are fitted into each branch to approximate the size. Furthermore, the shape and size of the tree cross-sections can be more accurately analyzed using, e.g., free-form curves (Pfeifer and Winterhalder, 2004).

In this paper we propose a new method for automatically approximating the size parameters and structures of trees from point clouds. The basic assumptions of the method are that the point cloud is a sample of a surface in the 3D-space and the surface, i.e. the tree, can be locally approximated with cylinders. Other a priori assumptions about the data and structure of trees are used as well. The basis of the method is a local approach where the point cloud is covered with small neighborhoods which conform to the surface. Then these neighborhoods are geometrically characterized and, based on these characterizations, the neighborhoods are classified into trunk, branch, and other points. Finally, cylinders are fitted to proper subsets to approximate the size. Notice that voxel spaces are not used, although partitions of the point cloud into cubical cells are used to produce the coverings quickly.

2 OBTAINING THE POINT CLOUDS

In this paper we used two point cloud samples, one from a coniferous tree and the other from a broadleaf tree without leafs (see Fig. 1). The trees were scanned from three different directions to have a comprehensive cover of the branching structure. The scans were registered to a common coordinate system via spherical reference targets placed in the measured area. The measured point clouds for both trees have about 1.7 million points.

The coniferous tree was scanned with a Faro Photon120 and the broadleaf tree with a Leica HDS6100 terrestrial laser scanner.
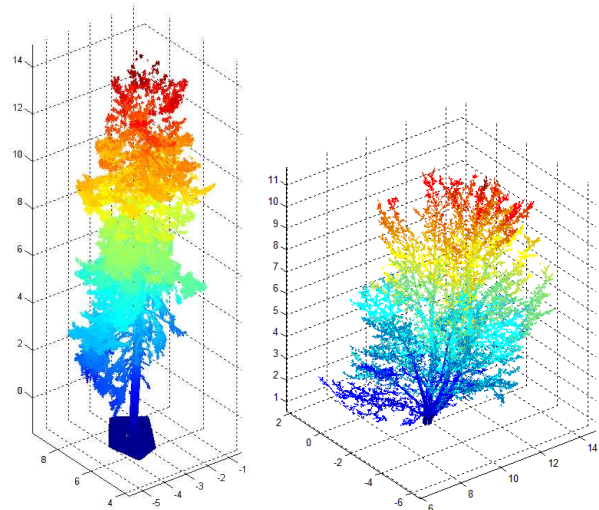

Figure 1: Point clouds from a coniferous (left) and a deciduous tree (right).

Both scanners use phase modulation with different carrier wavelengths for the distance measurement. The distance measurement accuracy is 2-3 mm at 25 m. Scanner specifications are given in Table 1.

	Faro Photon	Leica HDS6100
Wavelength	785 nm	650-690 nm
Unambiguity range	153 m	79 m
Field-of-view	$360° \times 320°$	$360° \times 310°$
Beam diameter	3.3 mm	3 mm
Beam divergence	0.16 mrad	0.22 mrad

Table 1: Scanner specification.

3 LOCAL APPROACH

The laser-scanner produced point cloud P_M from a tree is assumed to be a sample of a surface M embedded in the Euclidean space $\mathbb{R}^3$. The point cloud P_M inherits distances (metric) and neighborhoods (topology) from the embedding space $\mathbb{R}^3$. This metric locally approximates the intrinsic metric of the surface;

i.e. when the distances measured along the surface M are small, they are nearly equal to the distances of the embedding space which are measured 'directly through the tree and air'. Furthermore, the structure and size of trees can be locally reasonably approximated with cylinders. This suggests that although the global detailed structure of the surface M is complex and unknown, the local structure can be analyzed much more easily from the point cloud sample P_M. We propose the following local approach to the automatic size approximation:

1) Cover: The point cloud is covered with small neighborhoods.
2) Characterization: The neighborhoods are characterized geometrically.
3) Classification: Based on the characterizations the neighborhoods are classified to be part of trunk, branches, or other points.
4) Local size approximation: Determine proper subsets and fit cylinders to approximate the size.
5) Global characteristics: From local size data determine global characteristics such as the total volume and branch size distribution.

The details of these steps require approximations of the structures, such as metric and topology, of the surface M from the point cloud P_M. Next we give details on the approximation of some needed structures of M.

4 APPROXIMATION OF STRUCTURES

The point cloud $P_M \subset \mathbb{R}^3$ inherits the distance function d_P from the Euclidean embedding space $\mathbb{R}^3$ as a restriction of the Euclidean metric d of $\mathbb{R}^3$; i.e. $d_P(p,q) = d(p,q)$ holds for all $p, q \in P_M$. Similarly, the neighborhoods of P_M are defined as restrictions of the neighborhoods of $\mathbb{R}^3$ into P_M. A particularly important class of neighborhoods of P_M consists of the r-balls $B(p,r) = \{q \in P_M \mid d_P(p,q) < r\}$ (these are sometimes called fixed distance neighborhoods). Notice that these structures are globally defined for the point cloud, but they need not be good approximations of the corresponding global structures of the surface M. For example, the intrinsic metric d_M of M, which measures along the surface, gives always equal or greater distances than d_P, i.e. the inequality $d_P(p,q) \leq d_M(p,q)$ holds for all $p, q \in P_M$. Particularly, for points $p, q \in P_M$ that are in the tips of different but nearby branches, the distances in terms of metrics d_P and d_M satisfy $d_P(p,q) << d_M(p,q)$. However, locally the inherited structures of the point cloud are good approximations: for points $p, q \in P_M$ which are close to each other in terms of the intrinsic metric d_M, the approximation $d_M(p,q) \approx d_P(p,q)$ holds.

Next we show how connected components or connectedness of subsets of the surface M can be determined using a cover of M with r-balls. Two points of P_M are, by definition, in the same component if there exists a chain of small r-balls such that the consecutive balls are not disjoint (see Fig. 2). Notice that by this definition only those components of M whose distance measured with d_P is equal or larger than r can be recognized. Thus, the radius r of the balls should be so small that every r-ball itself is connected in the sense that it corresponds to a connected set of the surface M: The corresponding set of the r-ball $B(p,r) \subset P_M$ on M is defined to be the intersection of M and the r-ball of the embedding space $\mathbb{R}^3$ centered at point p, i.e. the set $\{q \in M \mid d(p,q) < r\}$, where d is the Euclidean metric of $\mathbb{R}^3$. Now if we have a covering C of P_M with small connected r-balls and know inside which balls each point is, we can determine the connected components with the following algorithm (see Fig. 3):

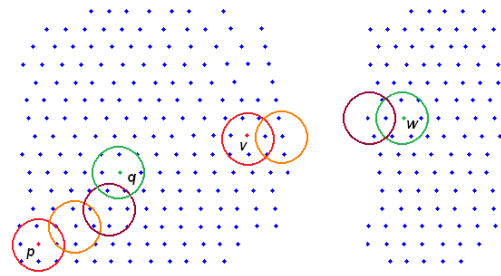1) Select any ball B of C which is not already assigned to some

Figure 2: Definition of connected components. The points p and q are in the same component because there is a chain of overlapping r-balls connecting them. Because there are no such chains between the points v and w, they are in different components.
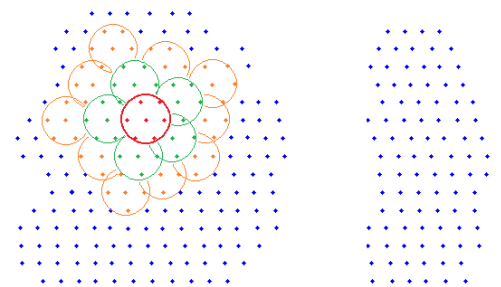

Figure 3: Determination of connected components. An arbitrary ball (red) is chosen and it is expanded by an iterative process where overlapping balls are added to the existing ones.

component and assign $A_0 = B$.
2) Find all the balls $B = \{B_j\}$ in the covering C such that $B_j \cap A_i$ is not empty and assign $A_{i+1} = A_i \cup B$.
3) If $A_{i+1} \neq A_i$ holds, repeat 2). If $A_{i+1} = A_i$ holds, then A_i is a component and continue from 1).

Finally, we show how the tangent planes and surface normals of the surface M for points of P_M can be approximated using the metric and neighborhoods defined above. In the approximation we fit planes to the neighborhoods using the total least squares method (Mitra et al., 2004). The solution of this problem is given by the eigenvectors of the scatter matrix $\mathbf{C}$ of the subset $B \subset P_M$ where the plane is fitted:

$$\mathbf{C}(B) = \Sigma_{i=1}^{m}(x_i - \bar{x})(x_i - \bar{x})^T, \qquad (1)$$

where $x_i \in B$, m is the number of points in B, and $\bar{x} = \frac{1}{m}\Sigma_{i=1}^{m}x_i$ is the mean of points x_i. The unit eigenvectors corresponding to the two largest eigenvalues of $\mathbf{C}$ span the tangent space and the unit eigenvector corresponding to the smallest eigenvalue is a surface normal.

With these approximated structures we can present details of our method outlined in the section 3. The first step in the method is to cover the point cloud with small sets.

5 COVER

We want to generate a cover $C = \{B_i\}$ of the point cloud P_M with r-balls B_i such that the center point of each ball is not included to other balls. The radius r should be as large as possible such that the local metric approximation $d_M \approx d_P$ still holds (virtually) everywhere and the r-balls are connected. A good
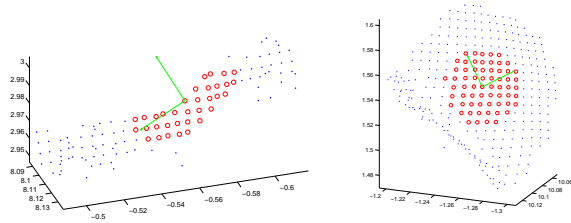
Figure 4: A subset B (red circles) is elongated (left) and this is reflected in the eigenvalues $(\lambda_1, \lambda_2, \lambda)$ and eigenvectors (u_1, u_2, u_3) of $\mathbf{C}(B)$ such that $\lambda_1 >> \lambda_2 \geq \lambda_3$ holds and u_1 approximates the direction B is elongated. When the eigenvalues λ_1 and λ_2 are approximately equal (right), there is no clear direction of elongatedness.

Figure 5: Classification of ground points. Left: The blue points denote the cover sets L_i that are initially classified as ground points ($r = 0.03$, $f_g = 100$, $p_g = 0.5$). Right: The final classification of ground points.

choice of radius r is about the size of the most branches. To generate the cover C quickly, we first partition the embedding space $\mathbb{R}^3$ into cubical cells such that the side length of the cubes is r. This partition of $\mathbb{R}^3$ also partitions the point cloud P_M into cubical cells and the cells where each point belongs can be directly calculated from the coordinates of the points. Now the ball $B(p, r)$ is contained in the r-cube containing the point p and in its 26 neighboring r-cubes. To generate the cover, we take the first point p_1 from P_M and calculate its r-ball neighborhood $B(p_1, r)$, which is the first ball in the cover C. Then we take another point from P_M which is not yet specified into a ball and define its r-ball. This process is continued until all points are specified into balls. During this process we also determine the balls each point belongs in.

The cover C can be used to generate another cover $C_L = \{L_i\}$ of P_M with larger sets L_i that are connected and conform to the surface M: each L_i is the union of all the balls $B_j \in C$ whose intersection with B_i is not empty (For example, in the Fig. 3 the set L_i corresponding to the red ball is the red ball together with the green balls). This process can be repeated to generate covers with even larger sets that are connected and conform to M.

6 CHARACTERIZATION

The second step in our local approach is the characterization of the cover sets B and L. The cover sets (and subsets of P_M in general) are geometrically characterized using the scatter matrix $\mathbf{C}(B)$. Assume that the points of B are approximately uniformly distributed over part of M. Furthermore, let $\{u_1, u_2, u_3\}$ be the unit eigenvectors of $\mathbf{C}(B)$ such that the corresponding eigenvalues satisfy $\lambda_1 \geq \lambda_2 \geq \lambda_3$. With these assumptions we now define some characteristics.

The unit vector u_1 approximates the direction in which the subset B is the most elongated (see Fig. 4). This gives our first characteristics, the *direction* $\mathbb{D}(B) = u_1$. On the other hand, how elongated the subset B is can be estimated by the ratio of λ_1 and λ_2 (see Fig. 4). Thus the second characteristics, the *elongatedness* $\mathbb{E}(B) = \frac{\lambda_1}{\lambda_2}$. Similarly with the *flatness* $\mathbb{F}(B) = \frac{\lambda_1}{\lambda_3}$ we can asses how planar the subset B is.

The vectors $\mathbb{D}(B)$ can be used, e.g,. to define the local direction of branches. Besides, because neighborhoods of same radius are usually more elongated for branch points than for trunk points, the value $\mathbb{E}(B)$ is usually larger for branch points than for trunk points. Similarly the flatness values for ground points are usually larger than for trunk points, which in turn usually have larger
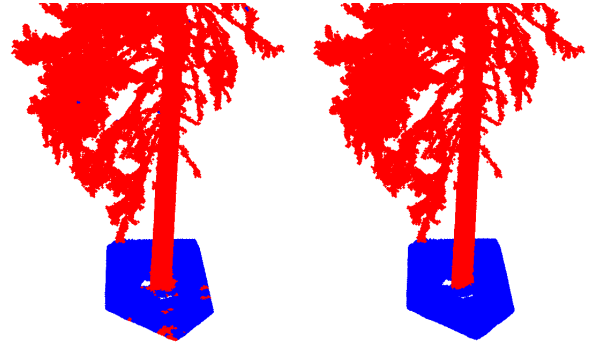
flatness values than branch points. A very large value of $\mathbb{F}(B)$ indicates that B is nearly planar.

The trunk is often nearly straight and has generally a different direction than the branches and the ground. For point clouds this means that the part of the surface M corresponding to the trunk has a 'characteristic' direction in the embedding space $\mathbb{R}^3$ given globally by a unit vector $V \in \mathbb{R}^3$. Then sets B that are *parallel with the trunk* direction have unit tangent vectors nearly parallel to V; i.e. the value $\mathbb{P}_V(B)$, defined as the maximum value of the dot product $V \cdot v$, where v is a unit tangent vector of B, is close to one. The characteristic direction vector V of a trunk can be often be known a priori quite accurately or it can be approximated well with the vector $\mathbb{D}(P_M)$ (if the tree is taller than it is wide).

7 CLASSIFICATION

When the cover sets of C and C_L are geometrically characterized, the next step in our method is to use these characterizations to classify the sets as ground (other), trunk, and branch points. We give examples how to employ the characterizations together with a priori assumptions on data and the structure of the trees.

Often the ground points have the most planar neighborhoods in the point clouds, and they are not parallel to the trunk. Using this assumption we take a high flatness value f_g and small parallelism value p_g and find the part $G_0 = \{L_i \,|\, \mathbb{F}(L_i) > f_g \,\&\, \mathbb{P}_V(L_i) < p_g\}$ of the cover C_L (see Fig. 5). Then we either take the mean of points G_0 or the largest connected component of G_0, which is near the true 'ground level'. The mean of G_0 is a faster way to approximate the ground level, but the largest connected component of G_0 is more reliable. Then the ground points G can be defined as all those sets L_i which are near the ground level and are not parallel to the trunk, i.e. $\mathbb{P}_V(L_i)$ is small (see Fig. 5).

Next we give an example of the initial classification of the trunk points. They are those cover sets $T_0 = \{L_i\}$ that are not classified as ground points and which are 1) parallel to the trunk direction V, i.e. $\mathbb{P}_V(L_i) > p_t$ holds, and 2) not elongated, i.e. $\mathbb{E}(L_i) < e_t$ holds, and 3) quite flat, i.e. $\mathbb{F}(L_i) > f_t$ holds. The numbers p_t, e_t, and f_t should be such that they will not exclude any or only a few real trunk point but at the same time will exclude most of the real branch points. By trial and error we have determined suitable parameter values, see Fig. 6. The suitable parameter values depend particularly on the radius r of the cover sets B. T_0 will probably contain numerous branch points but these are mostly separate from the trunk and contain only a few cover sets L_i. Therefore the final classification T can be defined
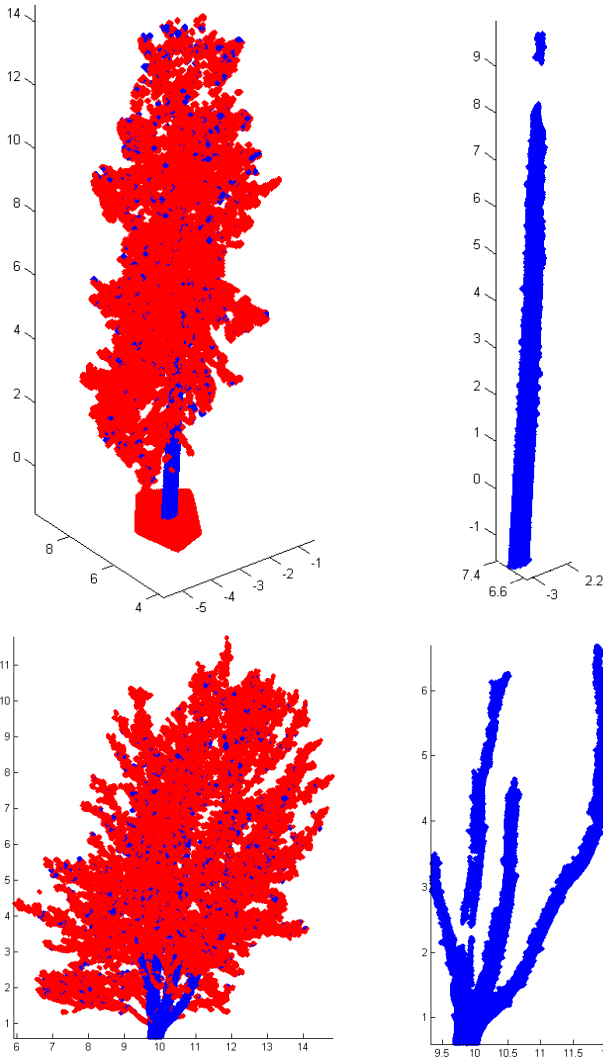
Figure 6: Classification of trunk points. Top: The blue points denote the cover sets that are initially classified as trunk points ($r = 0.03$, $p_t = 0.8$, $e_t = 3$, $f_t = 15$). Bottom: The final classification of trunk points, where only the largest components (over 2000 points) of the initial classification are accepted.

by excluding those connected components of T_0 that have a small number of points (see Fig. 6). On the other hand, if the trunk is approximately straight, we can have an estimate of the trunk axis from the largest component of T_0 and then define T by excluding all the small components of T_0 that are far enough from the axis.

The balls in the cover C that are not classified either as ground or trunk points are classified as the initial set Br_0 of branch points (see Fig. 7). The final set of branch points can be Br_0 but there are small components that are mostly useless for further analysis. Therefore the final set Br can be defined by excluding all the small components of Br_0.

8 LOCAL SIZE APPROXIMATION

The fourth step in the local approach is to approximate the size of trunk and branches by fitting cylinders into proper subsets of P_M. At first the separate components of the trunk and branches are determined and the components are studied one at a time. The
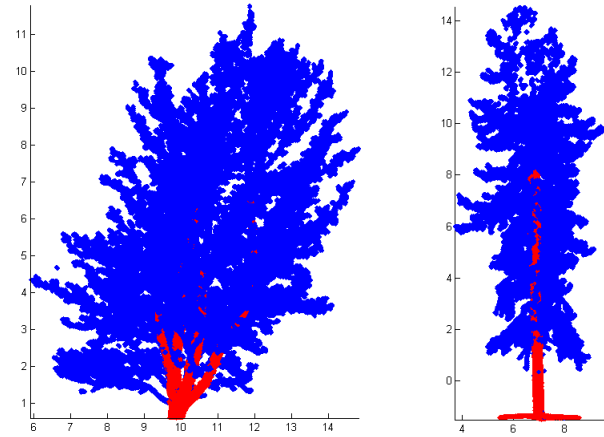

Figure 7: Classification of branch points.

proper subset of each component where to fit cylinders, is determined by the covers C and C_L and geometric characterizations such as those defined in section 6.

Trunk or branch components that have no bifurcations and whose radius is large compared to the cover sets can be analyzed using the so-called cylinder following (Pfeifer et al., 2004). Next we discuss how to analyze general components with bifurcations. At first we choose one of the cover sets L_0 of the connected component. Then L_0 is expanded into the set S, which is used for cylinder fitting, by its neighboring cover sets and it is expanded until it is round the branch or trunk. Furthermore, we need to make sure that S is not too much curved or has no bifurcations but is approximately straight. For this we can use the set directions $\mathbb{D}(B_i)$ of the r-balls B_i in S and exclude those balls whose direction deviates too much from most of the balls. We should use the direction vectors of the corresponding larger sets L_i for those balls B_i that are not clearly elongated (small $\mathbb{E}(B_i)$ values). In addition, there should be some upper length for S which can stop the expansion of L_0. Finally, there may be some other criteria for S. For example, S should have enough points and it should be round enough, which can be assessed from the ratio $\frac{\lambda_2}{\lambda_3}$ of the two smallest eigenvalues of $\mathbf{C}(S)$: the ratio indicates the ratio of the largest and the smallest extent of the set S when projected to the plane orthogonal to $\mathbb{D}(S)$. Thus the ratio value close to one is an indication of the roundness of cylinderlike sets. When an acceptable subset S is formed, the process of dividing the component into proper subsets is continued, preferably at neighboring cover sets of S, until all the cover set are dealt with.

For each acceptable subset S, we fit a cylinder using the total least square fitting (Lukacs et al., 1998). Because the problem is non-linear (Madsen et al., 2004), we need good initial guesses for the cylinder parameters which are the axis direction vector D_0, the position vector of an axis point P_0, and the radius R_0. Good guesses are $D_0 = \mathbb{D}(S)$, $P_0 = \bar{S}$ (the mean of the points S), and $R_0 = \sqrt{\lambda_2}$, where λ_2 is the second largest eigenvalue of $\mathbf{C}(S)$ corresponding to unit eigenvectors. Error terms such as the standard deviation and the distances of the points from the fitted cylinder can be easily calculated. Depending on the error terms, the result of the fitting is accepted or rejected. One possibility is to do another fitting where those points that had a large distance from the first fitted cylinder are excluded, and possibly use the results of the first fitting as new initial guesses. The length of the cylinder can be calculated by projecting all the points of S into the cylinder axis using the standard dot product.

For parts of trunk where there are lots of points round the trunk,
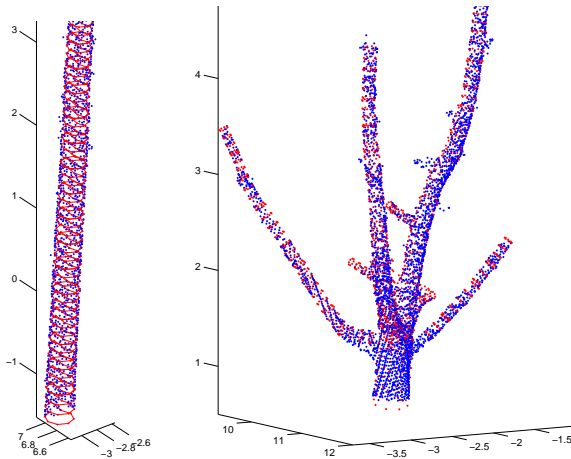
Figure 8: Cylinder fitting for trunk. The point clouds (blue points) are thinned out.
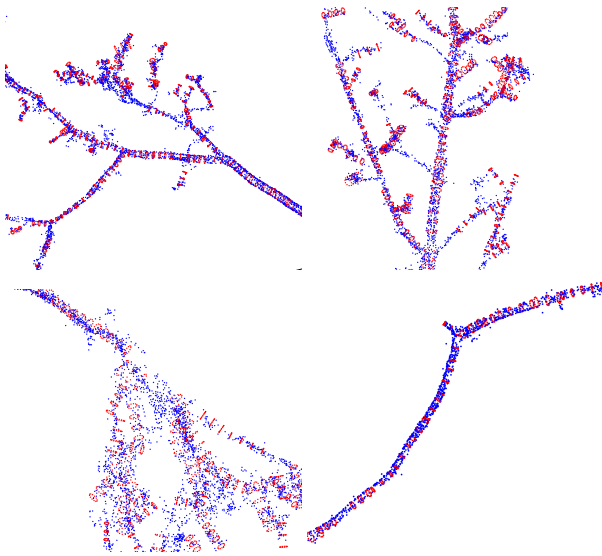

Figure 9: Cylinder fitting for branches. The upper branches are from deciduous tree and the bottom branches are from coniferous tree. The point clouds (blue points) are thinned out.

the fitting gives good approximation, see Fig. 8. For branches the fitting is usually less good because there are much less points that, furthermore, usually are not round the branch but cover only one side of the branch. In addition, the 'noise level' is larger compared to the size (see Fig. 9).

9 RESULTS AND DISCUSSION

The final step in our method is to approximate the total above-ground volume and branch size distribution from the cylinder fitting data (radius and height). In our examples, the calculated volume approximations are 0.26 and 0.23 cubic meters for the coniferous and the deciduous tree, respectively. These are probably underestimations because parts of the trunk and branches are not covered in the point clouds. Furthermore, there are also lot of parts that are poorly covered by the point clouds and whose size we could not estimate. This was particularly true for the upper parts of the trees and for small branches (see the parts without fitted cylinders in Fig. 9).

The calculated branch size distributions, i.e., the length and vol-
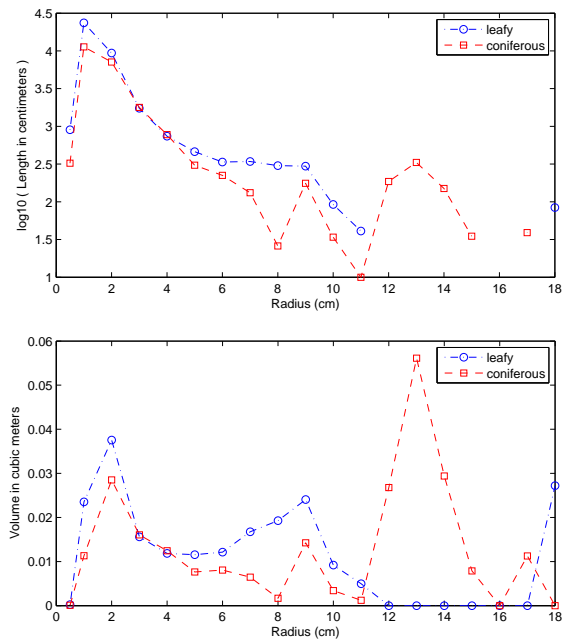

Figure 10: Distributions of branch length (above) and volume (below) as functions of branch radius.

ume of the branches as functions of the branch radius, are shown in Fig. 10 (the trunks are also included in the distributions). The distributions show, as expected, that most segments, in terms of length, are in the branches of the smallest radii whereas most of the volume is in the trunk segments. There were no direct measurements of the branch sizes so the accuracy can be assessed only based on the visual inspection of the point cloud and fitted cylinders. The accuracy seems to be good for trunk parts and for thick branches that have many points. On the other hand, as expected, for the smallest branches the accuracy is the poorest and approximations can have large relative errors. Particularly problematic are the small branches of coniferous trees because of their needles. Deciduous trees probably have the same problem when there are leaves (our scan was from a deciduous tree without its leaves). However, the approximation errors of small branches may not be a problem if we are only interested in a size resolution of some centimeters. Furthermore, although the absolute distribution values may have quite large errors, the relative distribution values are probably more accurate, particularly if there are systematic errors that can be estimated fairly well.

10 CONCLUSIONS AND FUTURE WORK

In this paper we have presented a new method to automatically approximate the size and structure of single trees from terrestrial laser scanner produced point clouds. A particular aim has been the approximation of branch size distribution which is essential in carbon cycle estimations. The general constitutive assumption of the method is that the point clouds are specimens from surfaces embedded in the 3D Euclidean space. Additionally, specifically for the application, the surfaces are supposed to be locally approximately like cylinders. The main steps of the procedure are covers of the point cloud with small sets, the classification of the cover sets based on their geometric characterizations, and cylinder fitting for size approximation. The method is very general and can be applied with minor changes to other applications as well: The coverings and their geometric characterizations are directly applicable to general point clouds and the classification of cover

sets is easy to modify for other applications where separation and recognition are important.

The method was demonstrated by computing approximations of the total above-ground volumes and branch size distributions of a deciduous tree and a coniferous tree. However, these demonstrations are only the beginning of future work in which the implementation of the method is developed further. Furthermore, the reliability and accuracy of the method and its realizations has to be analyzed. This includes, for example, the classification of the cover sets and how the classification depends on the size of the cover sets and the parameter values chosen for the geometric characterizations. The accuracy of the size approximations, particularly for small branches with needles and leaves, must also be evaluated. Moreover, the dependency of the size approximations and the classification on the number of points needs to be assessed. Also, the dependency of the total volume and branch size distribution approximations on the cover extent (the number of different scan positions) should be studied. The method and its implementations, in their many aspects, should also be compared with various other existing approaches.

ACKNOWLEDGEMENTS

REFERENCES

Gorte, B. G. H., Pfeifer, N. (2004). Structuring laser scanned trees using 3D mathematical morphology. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35: 929-933.

Lukacs, G., Martin, R., Marshall, D. (1998). Faithful least-squares fitting of spheres, cylinders and tori for reliable segmentation. *Proceedings of the 5th European Conference on Computer Vision*, 671-686.

Maas, H. -G., Bienert, A., Scheller, S., Keane, E. (2008). Automatic forest inventory parameter determination from terrestrial laser scanner data. *International Journal of Remote Sensing*, 29(5): 1579-1593.

Madsen, K., Nielsen, H. B., Tingleff, O. (2004). *Methods for Non-Linear Least Squares Problems (2nd ed.)*. Informatics and Mathematical Modelling, Technical University of Denmark, 60 pages.

Mitra, N. J., Nguyen, A., Guibas, L. (2004). Estimating Surface Normals in Noisy Point Cloud Data. *International Journal of Computational Geometry and Applications*, 14(4-5): 261-276.

Pfeifer, N., Gorte, B., Winterhalder, D. (2004). Automatic Reconstruction of Single Trees from Terrestrial Laser Scanner Data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Istanbul, Turkey, 35(B5): 114-119.

Pfeifer, N., Winterhalder, D. (2004). Modelling of tree cross sections from terrestrial laser-scanning data with free-form curves. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36: 76-81.

Rutzinger, M., Pratihast, A. K., Oude Elberink, S., Vosselman, G. (2010). Detection and modelling of 3D trees from mobile laser scanning data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(5): 520-525.