



Digital Analytics and Robotics for Sustainable Forestry

CL4-2021-DIGITAL-EMERGING-01

Grant agreement no: 101070405

DELIVERABLE 4.3

Semantic Segmentation System for Forest Environments

Due date: Month 24 (August 2024)

Deliverable type: R

Lead beneficiary: UBONN

Dissemination Level: PUBLIC

Main author: Meher V. R. Malladi, Tiziano Guadagnino, Jens Behley, Cyrill Stachniss

1 Introduction

Geometric models of the world represent the environments using 3D points, surfaces, or similar geometric primitives but do not provide an interpretation what kind of entities have been observed and these models provide no semantic meaning. The objective of this deliverable is to add semantic knowledge to the data acquired by the mobile mapping system developed in the project using modern machine learning methods such that objects (e.g., trees) can be identified, separated, counted, inspected, etc. This semantic information adds the required information needed for downstream tasks, such as deriving tree traits, which are relevant for providing a decision support system for forestry management.

In this deliverable, we summarize the developments to achieve a semantic interpretation of the LiDAR data acquired by the mobile mapping systems in real forests. We aim at computing a semantic label for each measured point from the acquired LiDAR data to yield a panoptic segmentation [1] of forest scenes, specifically, to semantically segment ground, shrub and trees along with instance information of trees.

In contrast to structured urban environments, where LiDAR perception has already seen promising results for panoptic segmentation [2], [3], the forestry environment poses novel challenges caused by the unstructured nature and organic growth of the forests. In particular, single scan LiDAR measurements usually cover only a very limited area due to the occlusion caused by the dense vegetation. At the same time, the LiDAR measurements are also more affected by noise due to the sparsity of tree foliage. This renders LiDAR perception in forestry environments a challenging task.

Realistic, domain-specific datasets are often a crucial foundation for developing innovative solutions and approaches as they provide novel challenges that cannot be efficiently solved by off-the-shelf solutions [4]. However, despite increasing interest in forestry robotics, only a few real-world robotics datasets and benchmarks are available in this important domain. This makes the development of new techniques and the comparison to existing ones difficult. At the outset of the DigiForest project, there was a particular lack of labeled data, and this is a challenge we address through Deliverable D7.2 of the DigiForest project, where we provide a longitudinal dataset for forestry robotics with semantic annotations. This dataset enables developing deep-learning based panoptic segmentation approaches for forestry environments.

Simultaneous to our work on the Deliverable D7.2 dataset, we first developed a geometric pipeline to extract tree instances from aggregated LiDAR scans [5]. Using the tree instance information, tree-specific traits like the so-called diameter at breast height (DBH) can be studied. Our geometric approach avoids the need of labeled data being available, and, with a limited number of parameters, is simple to tune and adapt. This approach remains useful in scenarios where labeled data is limited. However, our geometric method lacks the capability to semantically segment the scene. Therefore, we then leveraged the Deliverable D7.2 dataset we produced and developed a deep learning-based approach to provide full point-wise semantic and instance information of the whole forest scene.

More specifically, for getting a complete semantic interpretation of the whole scanned environment, i.e., plot-level information, we learn a panoptic segmentation model that allows us to directly predict a semantic class for each point of the LiDAR data, but also derive instance information for the trees. We implemented an approach that exploits a sparse voxel-based architecture using the MikowskiEngine [6] to extract rich features from the aggregated point cloud to estimate the semantics and instance information. Building on this panoptic segmentation of the forest scene, we then showcase the estimation of tree DBH, using the segmentation results as a foundation,

aligning with Task T5.1. This semantic information underpins precise forest data for informed decision-making, as outlined in Tasks T5.1 and T5.3. The output will be integrated with the map server (as described in D4.2) and will be accessible through a multi-layer semantic map, which enables other modules to use it for downstream analysis.

2 Semantic Interpretation in Forestry Environment

We developed two approaches for semantic interpretation of forests: (i) a geometric pipeline focusing on tree instance segmentation and (ii) a learning-based approach to derive a complete panoptic segmentation of the input LiDAR data.

We initially developed a geometric approach to gain practical experience with processing the provided LiDAR point clouds and to gain insights into challenges for a traditional pipeline for tree instance segmentation. This approach provided valuable insights and proved effective in scenarios where deep learning models are impractical due to insufficient labeled data. However, recognizing its limitations, particularly in separating tree instances from shrubs, we developed a learning-based approach leveraging a labeled forestry (c.f. Deliverable D7.2) dataset that we simultaneously provide. This method offers a more fine-grained semantic interpretation, allowing us to identify other semantic classes such as ground and shrub, and refine tree segmentation to distinguish between stem and foliage.

In the following, we first detail the geometric segmentation approach and then describe our learning-based approach.

2.1 Geometric Tree Instance Segmentation

Our geometric segmentation pipeline is designed to work on raw aggregated point clouds from LiDAR sensors, focusing on the problem of tree instance segmentation. We provide an overview of our pipeline in Fig. 1. Our open-source implementation is available at: https://github.com/PRBonn/forest_inventory_pipeline.

The terrain in forests shows significant variations in height and contains substantial under-canopy vegetation. Our geometric segmentation approach considers no semantics and is aimed solely at identifying trees. We preprocess an input point cloud with the aim of filtering out the ground, bushes, and any small near-ground structures. We first minimally denoise the cloud and apply the cloth simulation algorithm proposed by Zhang et al. [7] to compute a ground segmentation. Their method inverts the z -axis of the point cloud \mathcal{P} and simulates the interaction of a rigid cloth covering the inverted ground surface, extracting the set of ground points \mathcal{P}^G .

For points $\mathbf{p} = [p_x, p_y, p_z]^\top \in \mathcal{P}$ and $\mathbf{p}^i \in \mathcal{P}^G$, we interpolate the ground elevation of a point $h(\mathbf{p})$ as

$$h(\mathbf{p}) = \frac{\sum_{\mathbf{p}^i \in \mathcal{N}} w(\mathbf{p}, \mathbf{p}^i) p_z^i}{\sum_{\mathbf{p}^i \in \mathcal{N}} w(\mathbf{p}, \mathbf{p}^i)} \quad (1)$$

$$w(\mathbf{p}, \mathbf{p}^i) = d_{xy}(\mathbf{p}, \mathbf{p}^i)^{-2}, \quad (2)$$

where $d_{xy}(\mathbf{p}, \mathbf{p}^i)$ is the L2 norm over the x-y coordinates of the points, which is also used to define the local neighborhood \mathcal{N} around a point \mathbf{p} . We subtract $h(\mathbf{p})$ from p_z of each point, producing a height normalized point cloud. We then clip out points below 1 m in z -height to remove undergrowth vegetation. Examples of the

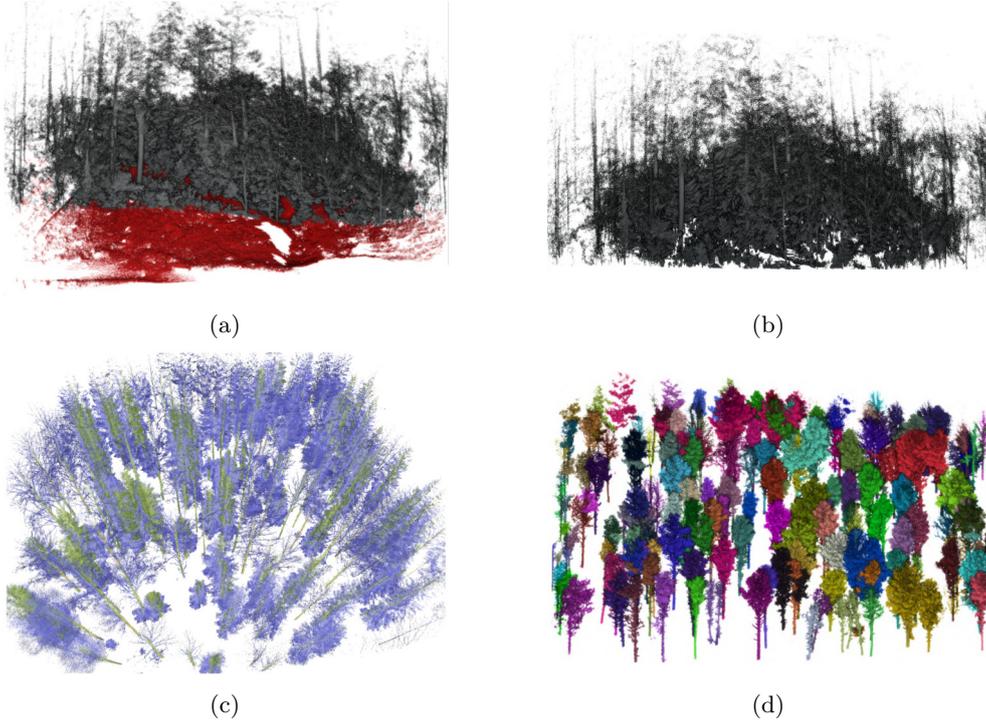


Figure 1: The results of different stages of our geometric tree instance segmentation pipeline are shown in (a) - (d). In (a), from an input LiDAR point cloud, the ground is first segmented and shown in red. In (b), the point cloud is height-normalized using the ground information. In (c), dense parts of the cloud are identified as cluster cores, guiding the tree instance segmentation. In (d), we show a result of our tree instance segmentation.

ground segmentation and height normalization results are shown in Fig. 1a and Fig. 1b respectively.

As the core of our segmentation approach, we apply Quickshift++ [8], a modification of the original Quickshift density-based clustering algorithm [9]. Following is a brief summary of Quickshift++ while illustrating how we use it in the context of our problem.

Let $r_k(\mathbf{p})$ for a point $\mathbf{p} \in \mathcal{P}$ be the distance of \mathbf{p} to its k -th nearest neighbor. For the true density $f(\mathbf{p})$ of a point \mathbf{p} , the k -NN density estimate of it is defined as

$$f_k(\mathbf{p}) = \frac{k}{n v r_k(\mathbf{p})^3}, \quad (3)$$

where n is the number of points in \mathcal{P} and v is the volume of a unit ball in \mathbb{R}^3 . In the first step, Quickshift++ aims at identifying the modes of $f(\mathbf{p})$ as cluster cores. These modes represent regions of locally high density. We define the mutual k -NN graph $G(\lambda)$ for a density level λ with vertices \mathcal{V} and edges \mathcal{E} as

$$\mathcal{V} = \{\mathbf{p} \in \mathcal{P} \mid f_k(\mathbf{p}) \geq \lambda\} \quad (4)$$

$$\mathcal{E} = \{\{\mathbf{p}_i, \mathbf{p}_j\} \mid \|\mathbf{p}_i - \mathbf{p}_j\| \leq \min(r_k(\mathbf{p}_i), r_k(\mathbf{p}_j))\}. \quad (5)$$

The connected components of $G(\lambda)$ approximate the connected components of the λ -level sets of the true density $f(\mathbf{p})$. If the λ -levels of $G(\lambda)$ are scanned top-down, every

new distinct connected component that appears corresponds to a local maximum at approximately density level λ . By then taking the corresponding connected components instead in $G((1 - \beta)\lambda)$, with the $(1 - \beta)$ multiplicative factor allowing for fluctuations in local density, we can identify cluster cores as the regions surrounding local modes.

In our pipeline, we take a top-down approach to tree segmentation. We first project the height-normalized cloud onto the XY-plane. We then identify the cluster cores, reasoning that the trunks should primarily correspond to the densest regions. These cluster cores are then unprojected back to their corresponding 3D points. Subsequently, the remaining points are clustered via the second step of Quickshift++ wherein each point is moved closer to its nearest neighbor with the highest density $f_k(\mathbf{p})$ until it reaches some cluster core. An example of how this density based clustering looks like can be seen in Fig. 1c, where the green regions correspond to the cluster cores or regions of high density and the purple regions are points with low density. The two main parameters of the approach are k for the k -th nearest neighbor and β for the fluctuation in density allowed in a cluster. The final result of how our geometric tree instance segmentation looks like is shown in Fig. 1d.

2.2 Forest Panoptic Segmentation

Our approach to panoptic segmentation builds upon recent advancements in deep learning which have demonstrated significant potential in semantic and instance segmentation in urban and agricultural environments. However, applying these techniques directly to forestry presents unique challenges due to the hierarchical and overlapping nature of the labels, especially for the tree, stem and canopy classes as in the dataset we produced for Deliverable D7.2. To effectively leverage deep learning for forest scene interpretation, it is necessary to adapt these methods to handle the specific structure and requirements of our dataset.

To address these challenges, we developed a panoptic segmentation approach tailored to forest scenes captured by LiDAR sensors. As we work with point clouds, we use Minkowski Engine [6], an auto-differentiation library for sparse tensors suitable for efficiently processing sparse 3D point clouds. The Minkowski Engine utilizes sparse tensor representations, which are crucial for handling the large, mostly empty spaces typical of LiDAR data, by only computing outputs for non-empty regions. This leads to significant savings in both memory and computational resources compared to dense convolutional neural networks.

We take a LiDAR point cloud as input to our model and generate a semantic and instance segmentation. Our network architecture is shown in Fig. 2. We use as a backend the MinkUNet architecture [6], which is an extension of the U-Net model specifically optimized for sparse data using the Minkowski Engine. Our backend has 785k trainable parameters and produces a feature embedding vector of size 32 for each point in the input LiDAR scan. The encoder and decoder in the backend use a rectified linear unit (ReLU) activation function throughout. The feature embedding is then passed on to two network heads in parallel: a semantic segmentation head and an offset-vector prediction head. The semantic segmentation head is a two layer multi-layer perceptron (MLP). It predicts per-point classes as ground, shrub, stem, or canopy and has an output depth equal to the number of semantic classes and a softmax activation function. It is trained with the standard cross-entropy loss, denoted as \mathcal{L}_{sem} given by

$$\mathcal{L}_{\text{sem}} = \frac{1}{N} \sum_{i=1}^N \text{CE}(s_i, s_i^*), \quad (6)$$

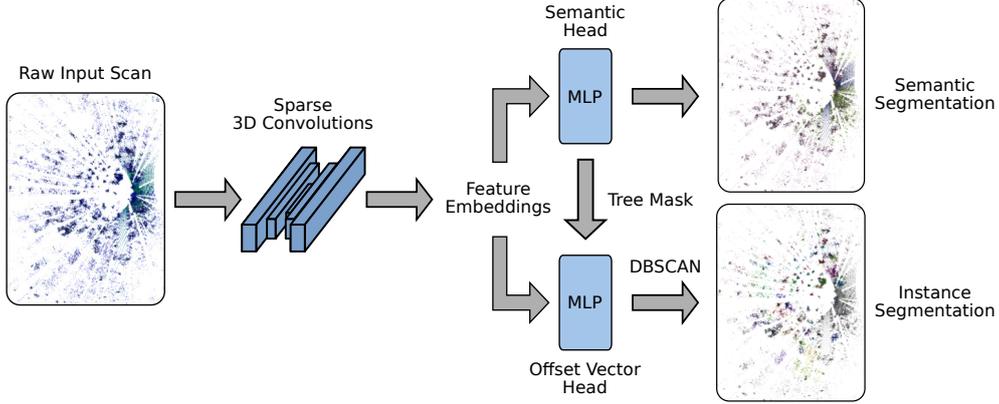


Figure 2: Overview of our panoptic segmentation pipeline. We take a raw LiDAR scan as input to a MinkUNet backend which uses the Minkowski Engine for sparse 3D convolutions and generate a feature embedding per LiDAR point. The embeddings are passed to two multi layer perceptron (MLP) heads resulting in a per point semantic and instance segmentation.

where N is the number of points, s is the predicted semantic class of a point and s^* is the semantic label.

The offset vector prediction head is a four layer MLP with an output depth of 3 as it predicts per-point offset vectors $\mathcal{O} = \{\mathbf{o}_1, \dots, \mathbf{o}_N\} \in \mathbb{R}^{N \times 3}$. These offsets represent the vector from each point to the geometric center of the corresponding tree instance. It is trained with an L1 regression loss [10] denoted as \mathcal{L}_{off} given by

$$\mathcal{L}_{\text{off}} = \frac{1}{\sum_{i=1}^N \mathbb{1}_{\{\mathbf{p}_i\}}} \sum_{i=1}^N \mathbb{1}_{\{\mathbf{p}_i\}} \|\mathbf{o}_i - \mathbf{o}_i^*\|_1, \quad (7)$$

where \mathbf{o}^* is the ground truth offset vector and $\mathbb{1}_{\{\mathbf{p}_i\}}$ is the indicator function indicating whether the point \mathbf{p}_i belongs to a tree instance. The combined loss \mathcal{L} is given by

$$\mathcal{L} = w_1 \cdot \mathcal{L}_{\text{sem}} + w_2 \cdot \mathcal{L}_{\text{off}}, \quad (8)$$

where w_i are the scalar weights for corresponding terms.

We use the AdamW [11] optimizer for training with a learning rate scheduler that reduces the learning rate starting from $1 \cdot 10^{-3}$ by a factor of 0.1 when the loss does not improve across ten epochs. During training, for the indicator function in \mathcal{L}_{off} we use a tree mask based on the shrub and canopy semantic labels to learn the offset vectors. During inference, we produce a tree mask from the points classified as stem or canopy. We apply the predicted offset vectors to these tree points, shifting them towards the estimated tree center. The shifted points are then clustered using the DBSCAN algorithm [12], grouping them into distinct tree instances. This method allows our network to perform both semantic segmentation and tree instance segmentation, addressing the unique hierarchical labeling structure present in our forestry dataset from Deliverable D7.2, as we show later in Sec. 3.2.

Method	Plot 1	Plot 2	Plot 3	Plot 4	Average
Silva et al. [13]	19.7	35.5	16.3	38.4	27.5
Dalponte et al. [14]	19.4	36.5	15.9	40.0	28.0
Li et al. [15]	39.0	48.5	34.1	53.5	43.7
Donager et al. [16]	61.2	56.1	33.0	48.5	49.7
Malladi et al. [5]	76.1	72.2	48.9	78.1	68.8

Table 1: Evaluation of geometric tree instance segmentation of different approaches on forest plots comparing panoptic quality (PQ).

3 Experimental Evaluation

We tested our approaches on LiDAR data acquired during data collection campaigns carried out for the DigiForest project. In subsequent sections, we summarize various experimental results, where we first show the performance of our geometric tree instance segmentation pipeline, then the deep-learning based panoptic segmentation approach, and finally a tree diameter at breast height estimation approach that builds upon our segmentation results.

3.1 Evaluation of Tree Instance Segmentation

In this section, we briefly summarize the evaluation of our geometric tree instance segmentation approach. For further details on implementation details, we refer the reader to our associated publication (see Appendix A).

We carried out field campaigns and collected extensive data from forests near Stein am Rhein, Switzerland, and Evo, Finland during March and May 2023 respectively. LiDAR sensor rigs included a Hesai Pandar XT-32, Leica’s BLK sensors, and the ANYmal robotic platform mounted with a Velodyne VLP-16. We show qualitative results on data from across these sensors and forests and focus our quantitative evaluation on the data collected with the Hesai sensor during the March 2023 Switzerland campaign.

We show comparisons against the methods of Silva et al. [13], Dalponte et al. [14], Li et al. [15], and Donager et al. [16], which are state-of-the-art geometric tree instance segmentation methods. To then evaluate the tree segmentation performance of the approaches, we use the standard panoptic quality (PQ) metric [1]. In our case, PQ_{tree} for tree segmentation is given by

$$PQ_{\text{tree}} = \frac{\sum_{(p,g) \in \text{TP}} \text{IoU}(p,g)}{|\text{TP}| + \frac{1}{2}|\text{FP}| + \frac{1}{2}|\text{FN}|} \quad (9)$$

where p, g represent the predicted and ground truth instance labels, TP is the set of true positives, FP is the set of false positives, FN is the set of false negatives, IoU is the intersection over union, and $|S|$ represents the cardinality of the set S . We follow the definition by Kirillov et al. [17] for data association of instance segments. Each predicted segment p is assigned to the ground truth segment g if its IoU is greater than 0.5 and is added to TP. Predicted instances p without a matching ground truth segment are added to FP and ground truth segments g without an associated prediction are added to FN.

Tab. 1 shows a quantitative comparison of our approach against the baselines. We see that our approach outperforms the baselines by a substantial margin. Furthermore,

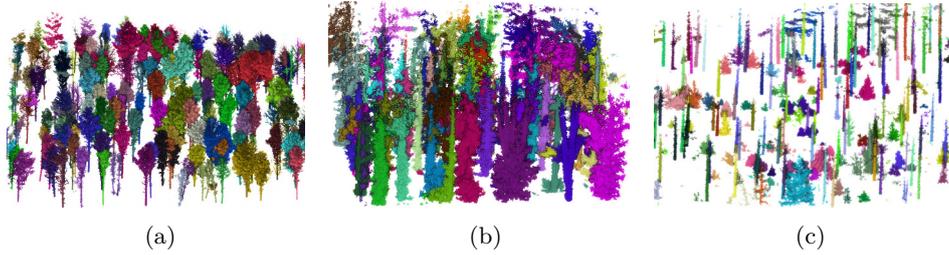


Figure 3: Qualitative results of our tree instance segmentation pipeline. (a) shows the instance segmentation for the Leica BLK sensor on the Evo site in Finland, (b) for the Hesai LiDAR in the Switzerland site, and (c) for the ANYmal platform equipped with a Velodyne VLP-16 in the Evo site.

in Fig. 3, we present the qualitative analysis of our pipeline showcasing the robustness of our approach. Note that the data for these experiments was collected from several sensors and from different geographical locations, i.e., forests in Switzerland and Finland.

3.2 Results for Panoptic Segmentation

In this section, we evaluate the panoptic segmentation performance of our approach detailed in Sec. 2.2 and compare it to recent deep-learning based approaches.

We use the dataset from Deliverable D7.2 to both train and evaluate the various learning-based approaches we present here. This dataset consists of point-wise semantic annotations for ground, shrub, and tree classes. Each tree point is also further classified into either stem or canopy for a fine-grained distinction. The dataset also contains instance labels for trees. Hence, this dataset uniquely lends itself to investigating multiple different problems in the wider context of semantic interpretation of forest scenes using commercial rotating LiDARs.

We follow the data splits recommended in Deliverable D7.2 and have, in summary, six forest plots for training, three forest plots for validation, and three forest plots as test set. We tested a mix of recent approaches for semantic, instance and panoptic segmentation from both the wider automated driving domain and the more adjacent agricultural domain. Following established protocols [18]–[20], and given the mix of approaches we tested, where possible we report the standard intersection-over-union (IoU) metric for performance on semantic classes and mean IoU (mIoU) for overall semantic segmentation performance. For evaluating panoptic segmentation performance, we again use the panoptic quality metric [1], [21], where the per-class Panoptic Quality PQ_c is given by [21]

$$PQ_c = \begin{cases} \frac{\sum_{(p,g) \in TP} \text{IoU}(p,g)}{|\text{TP}| + \frac{1}{2}|\text{FP}| + \frac{1}{2}|\text{FN}|} & , \text{ if } c \text{ is thing} \\ \text{IoU}(p,g) & , \text{ if } c \text{ is stuff} \end{cases} \quad (10)$$

where compared to Eq. (9) for PQ_{tree} for the thing (tree) class, we now additionally handle the stuff classes (ground, shrub). For stuff classes, we follow the implementation by Behley et al. [22] and compute simply the IoU between all points p assigned to class c and the points g with class c in the ground truth. As defined in Eq. (10), this renders PQ_c for stuff classes the same as its IoU which allows easier comparison between purely semantic and panoptic segmentation approaches. The overall panoptic

Approach	IoU				mIoU	PQ _{tree}	PQ
	Ground	Shrub	Stem	Canopy			
Fusaro et al. [23]	79.4	66.0	42.6	12.8	50.2	-	-
Marks et al. [20]	-	-	-	-	-	46.7	-
Marcuzzi et al. [3]	65.9	53.9	-	-	-	57.0	58.9
Ours	79.5	72.7	80.8	48.2	70.31	58.4	70.2

Table 2: Results for semantic interpretation on the test set of the Deliverable D7.2 dataset. For stuff classes, IoU is equivalent to PQ_c as defined by Eq. (10). mIoU is computed considering the ground, shrub, stem, and canopy classes. PQ is computed considering the ground, shrub, and tree classes.

quality PQ over all classes \mathcal{C} , i.e., ground, tree, and shrub, is given by the average over class-wise panoptic qualities given by:

$$\text{PQ} = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \text{PQ}_c \quad (11)$$

We trained multiple off-the-shelf baselines for comparison and also our approach detailed in Sec. 2.2. For all approaches, we report the relevant metrics on the test set of Deliverable D7.2 dataset in Tab. 2. First, from the automated driving domain, we trained the range-image based approach by Fusaro et al. [23] as a semantic segmentation baseline. We used the ground, shrub, stem and canopy labels as the target semantic classes, and report the IoU and the mIoU across these classes. Then we trained the approach by Marks et al. [20] which showed promising results in the agricultural domain in instance segmenting leaves on plants. Here, we adapted it to perform instance segmentation of the trees. Similar to how they report PQ on the leaf class in their work, we report here the PQ on just the tree class. We then trained the transformer-based MaskPLS by Marcuzzi et al. [3] developed to perform panoptic segmentation in urban environments. MaskPLS predicts a set of non-overlapping binary masks, each representing a single instance belonging to either a thing or a stuff class. It cannot segment a tree instance also for stem and canopy simultaneously. Hence, we train it using ground, shrub and tree as target labels and report the IoU for ground and shrub, PQ_{tree} for tree, and mean PQ across the three classes as defined by Eq. (10) and Eq. (11).

However, our approach detailed in Sec. 2.2 for panoptic segmentation does not face this limitation. It is a purpose-designed forestry panoptic segmentation architecture capable of both fine-grained stem-canopy semantic segmentation and tree instance segmentation. We report the results of our approach in Tab. 2 and see that it outperforms all other baselines on semantic and panoptic segmentation metrics.

3.3 Results on Trait Estimation

The output of this deliverable, the semantic information of a forest scene, will be used in downstream tasks to develop a forest inventory (see Task T5.1) and decision support system (see Task T5.3). We show in this experiment the estimation of the so-called diameter at breast height (DBH) using the panoptic segmentation system from this deliverable, aligning with Task T5.1.

Tree DBH is a parameter widely considered the most crucial parameter in forestry [24]. Common approaches for DBH estimation perform either circle [25] or cylinder fit-



Figure 4: Qualitative results of our cylinder fitting approach, where fitted cylinders are shown in red. The point cloud has been cropped along its height to allow a better visualization.

Approach	RMSE [cm]
Krisanski et al. [27]	6.08
Ours + Cylinder Fitting	5.43
G.T. + Cylinder Fitting	3.69

Table 3: Results for DBH trait estimation on the Deliverable D7.2 dataset.

ting [26], [27], considering that either a manual segmentation of trees is available [26] or segmenting the trees as part of their approach as well [5], [27]. We report in Tab. 3 results of DBH estimation approaches on the forest inventory provided as part of the Deliverable D7.2 dataset. To evaluate an approach, we take each tree in the known ground truth and assign it to the closest predicted tree location using a nearest neighbor search on the tree position and a maximum search radius of 0.2 m. We then report the RMSE of DBH estimated of all trees in a plot averaged across all the plots in the forest inventory.

As a baseline, we report the results of the approach by Krisanski et al. [27]. They use a PointNet++ [28] model for semantic segmentation of trees and cluster short vertical slices of the tree points using HDBSCAN [29]. They then estimate the DBH using a RANSAC circle fit on the vertical slices. Next, we report the results of using our panoptic segmentation approach followed by our cylinder fitting pipeline from Malladi et al. [5]. The panoptic segmentation is used to obtain an instance segmentation of the trees while filtering for only tree stem points. A cylinder is then fit to each tree instance and the DBH of a tree is the diameter of the estimated cylinder. The accuracy of our approach is reported as “Ours + Cylinder Fitting” in Tab. 3. Finally, we use the ground truth semantic annotations directly to obtain the stem point cloud and then repeat the cylinder fitting methodology as above. The result of this is reported as “G.T. + Cylinder Fitting” in Tab. 3.

From the results, we see that our approach outperforms that by Krisanski et al. [27]. Comparing the results of using the ground truth annotations followed by cylinder

fitting, we can see that, as expected, the quality of stem segmentation itself influences the performance of DBH estimation. Better results can therefore be obtained by first improving the segmentation results in the pipeline. Furthermore, more sophisticated geometric primitives for estimating the DBH can be explored. We do also note that, from earlier studies [24], [30], the accuracy of DBH estimation using even terrestrial laser scanners typically falls within the range of 1 to 4 cm.

We also show in Fig. 4 a qualitative result of our approach to cylinder fitting on data collected in a forest plot in Evo, Finland.

4 Conclusion

In this document, we summarized our development of a semantic segmentation system for forestry environments. We developed two segmentation approaches, one for geometric tree instance segmentation and the other for learning-based forest panoptic segmentation. Both methods are suitable for forestry applications depending on the availability of data and the problem specifics. The output of this deliverable can then be used in downstream tasks to generate a holistic representation of the forest including tree instances. This can be used to derive relevant tree traits, such as tree diameter at breast height, to enable the forest decision support system to provide actionable information about the status of the forest.

References

- [1] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, “Panoptic Segmentation,” in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [2] A. Milioto, J. Behley, C. McCool, and C. Stachniss, “LiDAR Panoptic Segmentation for Autonomous Driving,” in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2020.
- [3] R. Marcuzzi, L. Nunes, L. Wiesmann, J. Behley, and C. Stachniss, “Mask-Based Panoptic LiDAR Segmentation for Autonomous Driving,” *IEEE Robotics and Automation Letters (RA-L)*, vol. 8, no. 2, pp. 1141–1148, 2023.
- [4] S. Stevens, J. Wu, M. J. Thompson, E. G. Campolongo, C. H. Song, D. E. Carlyn, L. Dong, W. M. Dahdul, C. Stewart, T. Berger-Wolf, W.-L. Chao, and Y. Su, “BioCLIP: A vision foundation model for the tree of life,” in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [5] M. Malladi, T. Guadagnino, L. Lobefaro, M. Mattamala, H. Griess, J. Schweier, N. Chebrolu, M. Fallon, J. Behley, and C. Stachniss, “Tree Instance Segmentation and Traits Estimation for Forestry Environments Exploiting LiDAR Data,” in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2024.
- [6] C. Choy, J. Gwak, and S. Savarese, “4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks,” in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [7] W. Zhang, J. Qi, P. Wan, H. Wang, D. Xie, X. Wang, and G. Yan, “An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation,” *Remote Sensing*, vol. 8, no. 6, 2016.
- [8] H. Jiang, J. Jang, and S. Kpotufe, “Quickshift++: Provably good initializations for sample-based mean shift,” in *Proc. of the Intl. Conf. on Machine Learning (ICML)*, 2018.
- [9] A. Vedaldi and S. Soatto, “Quick shift and kernel methods for mode seeking,” in *Proc. of the Europ. Conf. on Computer Vision (ECCV)*, 2008.
- [10] T. Vu, K. Kim, T. M. Luu, T. Nguyen, and C. D. Yoo, “SoftGroup for 3D Instance Segmentation on Point Clouds,” in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [11] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in *Proc. of the Intl. Conf. on Learning Representations (ICLR)*, 2019.
- [12] M. Ester, H. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proc. of the Conf. on Knowledge Discovery and Data Mining (KDD)*, 1996.
- [13] C. A. Silva, A. T. Hudak, L. A. Vierling, E. L. Loudermilk, J. J. O’Brien, J. K. Hiers, S. B. Jack, C. Gonzalez-Benecke, H. Lee, M. J. Falkowski, and A. Khosravipour, “Imputation of individual longleaf pine (*pinus palustris* mill.) tree attributes from field and lidar data,” *Canadian Journal of Remote Sensing*, vol. 42, no. 5, pp. 554–573, 2016.
- [14] M. Dalponte and D. A. Coomes, “Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data,” *Methods in Ecology and Evolution*, vol. 7, no. 10, pp. 1236–1245, 2016.

- [15] W. Li, Q. Guo, M. K. Jakubowski, and M. Kelly, "A new method for segmenting individual trees from the lidar point cloud," *Photogrammetric Engineering and Remote Sensing (PE&RS)*, vol. 78, no. 1, pp. 75–84, 2012.
- [16] J. J. Donager, A. J. S. Meador, and R. C. Blackburn, "Adjudicating perspectives on forest structure: How do airborne, terrestrial, and mobile lidar-derived estimates compare?" *Remote Sensing*, vol. 13, no. 12, 2021.
- [17] A. Kirillov, R. Girshick, K. He, and P. Dollar, "Panoptic Feature Pyramid Networks," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [18] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, J. Gall, and C. Stachniss, "Towards 3D LiDAR-based Semantic Scene Understanding of 3D Point Cloud Sequences: The SemanticKITTI Dataset," *Intl. Journal of Robotics Research (IJRR)*, vol. 40, no. 8–9, pp. 959–967, 2021.
- [19] K. Vidanapathirana, J. Knights, S. Hausler, M. Cox, M. Ramezani, J. Jooste, E. Griffiths, S. Mohamed, S. Sridharan, C. Fookes, and P. Moghadam, "WildScenes: A Benchmark for 2D and 3D Semantic Segmentation in Large-scale Natural Environments," *arXiv preprint*, vol. arXiv:2312.15364, 2023.
- [20] E. Marks, M. Sodano, F. Magistri, L. Wiesmann, D. Desai, R. Marcuzzi, J. Behley, and C. Stachniss, "High precision leaf instance segmentation for phenotyping in point clouds obtained under real field conditions," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4791–4798, 2023.
- [21] L. Porzi, S. R. Bulo, A. Colovic, and P. Kotschieder, "Seamless Scene Segmentation," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [22] J. Behley, A. Milioto, and C. Stachniss, "A Benchmark for LiDAR-Based Panoptic Segmentation Based on KITTI," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2021.
- [23] D. Fusaro, S. Mosco, E. Menegatti, and A. Pretto, "Exploiting local features and range images for small data real-time point cloud semantic segmentation," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2024.
- [24] X. Liang, V. Kankare, J. Hyypää, Y. Wang, A. Kukko, H. Haggrén, X. Yu, H. Kaartinen, A. Jaakkola, F. Guan, M. Holopainen, and M. Vastaranta, "Terrestrial laser scanning in forest inventories," *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*, vol. 115, pp. 63–77, 2016.
- [25] A. Proudman, M. Ramezani, S. T. Digumarti, N. Chebrolu, and M. Fallon, "Towards real-time forest inventory using handheld LiDAR," *Journal on Robotics and Autonomous Systems (RAS)*, vol. 157, p. 104240, 2022.
- [26] J.-F. Tremblay, M. Beland, R. Gagnon, F. Pomerleau, and P. Giguere, "Automatic three-dimensional mapping for tree diameter measurements in inventory operations," *Journal of Field Robotics (JFR)*, vol. 37, no. 8, pp. 1328–1346, 2020.
- [27] S. Krisanski, M. S. Taskhiri, S. G. Aracil, D. Herries, A. Muneri, M. B. Gurung, J. Montgomery, and P. Turner, "Forest structural complexity tool—an open source, fully-automated tool for measuring forest point clouds," *Remote Sensing*, vol. 13, no. 22, 2021.
- [28] C. Qi, K. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," in *Proc. of the Conf. on Neural Information Processing Systems (NeurIPS)*, 2017.

-
- [29] C. Malzer and M. Baum, “A hybrid approach to hierarchical density-based cluster selection,” in *Intl. Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2020.
- [30] D. Kükenbrink, M. Marty, R. Bösch, and C. Ginzler, “Benchmarking laser scanning and terrestrial photogrammetry to extract forest inventory parameters in a complex temperate forest,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 113, p. 102 999, 2022.

A Appendix

In this appendix, we include the full text of the associated publication [5].

M. Malladi, T. Guadagnino, L. Lobefaro, M. Mattamala, H. Griess, J. Schweier, N. Chebrolu, M. Fallon, J. Behley, and C. Stachniss, “Tree Instance Segmentation and Traits Estimation for Forestry Environments Exploiting LiDAR Data,” in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2024.

Tree Instance Segmentation and Traits Estimation for Forestry Environments Exploiting LiDAR Data Collected by Mobile Robots

Meher V. R. Malladi Tiziano Guadagnino Luca Lobefaro Matias Mattamala Holger Griess
Janine Schweier Nived Chebrolu Maurice Fallon Jens Behley Cyrill Stachniss

Abstract—Forests play a crucial role in our ecosystems, functioning as carbon sinks, climate stabilizers, biodiversity hubs, and sources of wood. By the very nature of their scale, monitoring and maintaining forests is a challenging task. Robotics in forestry can have the potential for substantial automation toward efficient and sustainable foresting practices. In this paper, we address the problem of automatically producing a forest inventory by exploiting LiDAR data collected by a mobile platform. To construct an inventory, we first extract tree instances from point clouds. Then, we process each instance to extract forestry inventory information. Our approach provides the per-tree geometric trait of “diameter at breast height” together with the individual tree locations in a plot. We validate our results against manual measurements collected by foresters during field trials. Our experiments show strong segmentation and tree trait estimation performance, underlining the potential for automating forestry services. Results furthermore show a superior performance compared to the popular baseline methods used in this domain.

I. INTRODUCTION

Forests are vital for our ecosystems [10]. They support critical processes like carbon sequestration and biodiversity in the biosphere, while also providing resources for timber industries and offering opportunities for human leisure in the anthroposphere [10], [27]. Monitoring and documenting the status of a forest are time-intensive tasks while the number of forests is decreasing in most countries. Robots, however, can perform extensive data collection in forests and may go in the future as far as realizing automated maintenance and even tree harvesting. Monitoring can yield precise information on tree count, species distribution, essential geometric traits like diameter at breast height (DBH), and more, constituting a forest inventory [20]. Foresters can use such detailed inventories to make accurate forecasts of stand growth, plan harvesting strategies, optimize species rotation cycles, and more, contributing to effective and sustainable foresting practices [29].

Meher V. R. Malladi, Tiziano Guadagnino, Luca Lobefaro, Jens Behley, and Cyrill Stachniss are with the Center for Robotics, University of Bonn, Germany. Cyrill Stachniss is additionally with the Department of Engineering Science at the University of Oxford, UK, and with the Lamarr Institute for Machine Learning and Artificial Intelligence, Germany. Matias Mattamala, Nived Chebrolu and Maurice Fallon are with the University of Oxford, UK. Holger Griess and Janine Schweier are with Swiss Federal Institute for Forest, Snow and Landscape Research, Switzerland.

This work has partially been funded by the European Union’s Horizon Europe research and innovation program under grant agreement No 101070405 (DigiForest), by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy, EXC-2070 – 390732324 – PhenoRob and by the Swiss State Secretariat for Education, Research and Innovation (SERI).



Fig. 1: Instance segmentation results on data collected with a mobile robotics platform in a forest near Evo, Finland. On the left, we show the ANYmal platform, which was used for data collection, and on the right the segmentation results of our approach on the collected data. Different colors of points indicate different instances.

Common sensor modalities for forest inventory include terrestrial laser scanners (TLS) [19] and UAV laser scanners [43]. TLS offers high resolution at the cost of limited spatial coverage, while UAV laser scanners have lower resolution but with larger coverage. Additionally, since UAV flights are usually performed above-canopy, occlusions can result in insufficient tree trunk detail. By using ground-based robotic platforms and below-canopy UAV flights, both problems of spatial coverage and trunk-canopy detail can be addressed. Although their resolution might not match TLS, they can still provide sufficient below-canopy detail and extended spatial extent.

This paper focuses on the problem of using laser scanner data collected with mobile robotic platforms to automate forest inventory: to obtain forest stand tree count and per-tree information like DBH. DBH, measured approximately 1.3 m above ground, is considered the most crucial parameter in forestry [20]. It is used in estimating forest biomass [23], developing forest growth models and more [21]. Hence, we developed our approach focusing first on DBH estimation. Within this context, the challenge involves two key aspects. Firstly, the instance segmentation of trees and secondly,

the subsequent modeling of these identified tree instances. Existing methods to address them predominantly work with modalities like airborne laser scanners [5] or TLS [6], [8], [42] and usually take a geometric approach. While modern deep-learning methods have seen good success in related fields [25], [33] and have recently been explored in forestry settings as well [9], [17], [26], they can suffer from problems of out-of-distribution error, differences in sensor platforms between training and testing, differences in forest structure and species distribution, and more. The difficulty is further compounded by the limited availability of open datasets, with the one from Hannah et al. [43] being one of few easily available datasets providing labeled terrestrial, aerial, and UAV laser scanner data. To the best of our knowledge, no labeled datasets from mobile laser scanning platforms are currently available.

The main contribution of this paper is a geometric tree instance segmentation and structural analysis pipeline for forestry data collected from mobile laser scanning platforms. Fig. 1 depicts an example of the type of data considered. Our geometric approach avoids the need for extensive data labeling, as would be needed for modern deep-learning models. With a limited number of parameters, our approach is simple to tune. The proposed method performs well when evaluated quantitatively for segmentation accuracy, DBH trait estimation, and shows good qualitative performance on data collected from varied sensor types. In sum, we make three key claims: Our approach is able to (i) segment trees from mobile laser scans of forest scenes with high segmentation accuracy and no or minimal labeling effort, (ii) fit geometric primitives to the data reliant on the results of the segmentation and achieve high accuracy on tree trait estimation, (iii) show a solid generalization performance across data from different sensor types. These claims are backed up by the paper and our experimental evaluation. The open-source implementation of our approach is available at: https://github.com/PRBonn/forest_inventory_pipeline.

II. RELATED WORK

Building maps of the environment from sensor data requires addressing the SLAM problem [11], [37] involving several subtasks such as incremental pose estimation [40], place recognition [41], loop closing [12], and optimization [2], [32]. Using laser scanning to map and study forests has been extensively studied [30], [44], with considerable attention given to airborne laser scanner [5], [18], [34], [38] and TLS data [1], [6], [8], [19], [20], [26], [42], [43]. Mobile laser scanning has only recently become a more viable approach, usually in the form of handheld or UAV platforms [3], [7], [22], [30]. For example, Proudman et al. [30] showcase a SLAM pipeline to enable real-time mapping of forests while simultaneously giving foresters insight into the captured tree structure. We aim to do similarly, to segment trees automatically and analyze traits while working with data collected from mobile robotic platforms, a first in literature to the best of our knowledge.

A standard approach for segmentation of airborne laser scanning data is that of Dalponte et al. [5] in which they rasterize a canopy height model derived from a point cloud, identify local maxima as treetops, and then use a region-growing algorithm to delineate trees. Silva et al. [36] take a similar approach while working with rasterized canopy height models. However, the rasterization can introduce approximation errors that degrades segmentation performance. In contrast, Li et al. [18] work directly on the point cloud, starting from the highest points and using a spacing threshold to iteratively assign points within a search radius to an instance. Both methods have been implemented as part of lidR [34], a popular package for airborne laser data analyses. Similarly, various approaches have been developed for TLS as well, but they critically rely on the high resolution of TLS [6], [8] or require manual interventions [1].

Proudman et al. [30] uses standard Euclidean clustering of locally aggregated mobile laser scans for an initial segmentation and a heuristics-based merge/discard scheme to find tree instances. They estimate DBH via a RANSAC-based circle fit. However, a more sophisticated segmentation approach is necessary when working with complex forest scenes, especially when thickly intertwining canopies exist. Circle fitting for DBH estimation is also done by Heo et al. [13], though they manually segment trees before their analysis. Most related to our approach is the one by Donager et al. [7], who also compare data collected from handheld platforms against TLS and airborne lasers for forest inventory. They show the potential of handheld sensors, reasoning that such platforms overcome the problem of occlusion apparent in TLS. From a slice of the point cloud around breast height, they first identify tree stems by thresholding rasters of geometric features and then clipping out regions. The DBH is estimated by RANSAC cylinder fitting through these points. The tree instances are then “grown” by converting the point cloud into a proximity graph and assigning points to identified stems based on the shortest path in this graph. We take a different approach to segmentation, exploiting the spatial density of points and similarly estimating DBH by cylinder fitting. In contrast to their work, we use a wider height range of points and refine cylinder fitting using a least-squares scheme.

Data-driven techniques have been applied to tackle forestry problems [9], [26], [35], [38], however, the limited availability of labeled point cloud datasets is to the detriment of such approaches. Only a few are openly available for TLS [19], [43], and to the best of our knowledge, none are available with data from mobile platforms. Krisanski et al. [17] used a PointNet++ [31] model for semantic segmentation trained on a mix of data from different sensor types. In a follow-up work [16], they extend their approach to calculate tree count and DBH estimates to be used for forest inventory. Once an input point cloud has been semantically segmented, they filter for stem-only points, cluster short vertical slices using HDBSCAN [24], and by treating these slices as cylinders estimate the radius using a RANSAC circle fit. Stem instances are obtained by merging cylinders

using geometric heuristics. They estimate the DBH by taking a mean of the cylinder radii between 1.0m and 1.6m above ground.

We propose in this paper a pipeline for geometric instance segmentation and DBH estimation of trees from robotic mobile laser scans, avoiding the need for extensive data labeling as it would be needed for a data-driven learning approach. We show improved performance on multiple baselines in our segmentation accuracy and DBH estimation evaluations. In conjunction with using mobile laser scans to cover large swathes of forest regions, our proposed pipeline to build forest inventories is a more affordable solution than manual alternatives with human operators and TLS.

III. OUR APPROACH TO FOREST INVENTORY

Our pipeline is designed to work on raw aggregated point clouds from LiDAR sensors and automate tree-level analysis to produce an accurate forest inventory. We first focus on the problems of tree instance segmentation and subsequently DBH measurement.

A. Preprocessing

The terrain in forests shows significant variations in height and contains substantial under-canopy vegetation. Our segmentation approach considers no semantics and is aimed solely at identifying trees. We preprocess an input point cloud with the aim of filtering out the ground, bushes, and any small near-ground structures. We first minimally denoise the cloud and apply the cloth simulation algorithm proposed by Zhang et al. [45] to compute a ground segmentation. Their method inverts the z-axis of the point cloud \mathcal{P} and simulates the interaction of a rigid cloth covering the inverted ground surface, extracting the set of ground points \mathcal{P}^G .

For points $\mathbf{p} = [p_x, p_y, p_z]^\top \in \mathcal{P}$ and $\mathbf{p}^i \in \mathcal{P}^G$, we interpolate the ground elevation of a point $h(\mathbf{p})$ as

$$h(\mathbf{p}) = \frac{\sum_{\mathbf{p}^i \in \mathcal{N}} w(\mathbf{p}, \mathbf{p}^i) p_z^i}{\sum_{\mathbf{p}^i \in \mathcal{N}} w(\mathbf{p}, \mathbf{p}^i)} \quad (1)$$

$$w(\mathbf{p}, \mathbf{p}^i) = d_{xy}(\mathbf{p}, \mathbf{p}^i)^{-2}, \quad (2)$$

where $d_{xy}(\mathbf{p}, \mathbf{p}^i)$ is the L2 norm over the x-y coordinates of the points, which is also used to define the local neighborhood \mathcal{N} around a point \mathbf{p} . We subtract $h(\mathbf{p})$ from p_z of each point, producing an elevation or height normalized point cloud. We then clip out points below 1m in z-height to remove undergrowth vegetation. An example result of the ground segmentation and height normalization is given in Fig. 2.

B. Tree Instance Segmentation

Following existing methods [1], [17], [30], we downsample the height-normalized cloud to remove duplicate points and speed up the segmentation. The size of voxelization is a parameter that can influence the upper bound of accuracy as desired. However, we note that the accuracy of DBH estimation typically falls within the range of 1 to 4 centimeters [20].

As the core of our segmentation approach, we apply Quickshift++ [14], a modification of the original Quickshift

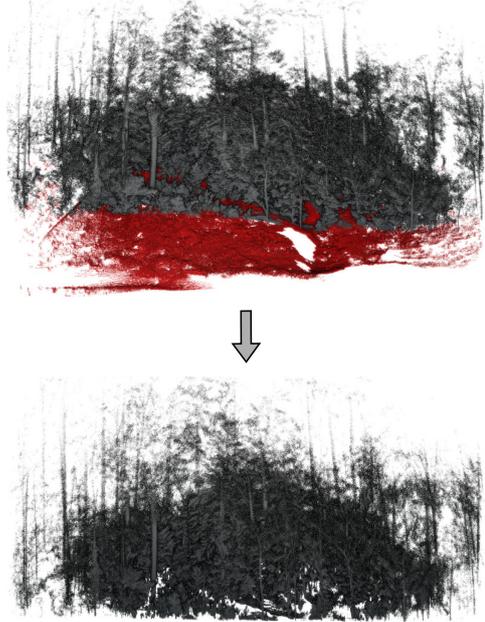


Fig. 2: Results of ground segmentation and height normalization steps. In the top image, points in red denote identified ground points. The ground segmentation is used to normalize the height, as shown in the image below.

density-based clustering algorithm [39]. Following is a brief summary of Quickshift++ while illustrating how we use it in the context of our problem. For more details, we refer the reader to the work by Jiang et al. [14].

Let $r_k(\mathbf{p})$ for a point $\mathbf{p} \in \mathcal{P}$ be the distance of \mathbf{p} to its k -th nearest neighbor. For the true density $f(\mathbf{p})$ of a point \mathbf{p} , the k -NN density estimate of it is defined as

$$f_k(\mathbf{p}) = \frac{k}{n v r_k(\mathbf{p})^3}, \quad (3)$$

where n is the number of points in \mathcal{P} and v is the volume of a unit ball in \mathbb{R}^3 . In the first step, Quickshift++ aims at identifying the modes of $f(\mathbf{p})$ as cluster cores. These modes represent regions of locally high density. We define the mutual k -NN graph $G(\lambda)$ for a density level λ with vertices \mathcal{V} and edges \mathcal{E} as

$$\mathcal{V} = \{\mathbf{p} \in \mathcal{P} \mid f_k(\mathbf{p}) \geq \lambda\} \quad (4)$$

$$\mathcal{E} = \{\{\mathbf{p}_i, \mathbf{p}_j\} \mid \|\mathbf{p}_i - \mathbf{p}_j\| \leq \min(r_k(\mathbf{p}_i), r_k(\mathbf{p}_j))\}. \quad (5)$$

The connected components of $G(\lambda)$ approximate the connected components of the λ -level sets of the true density $f(\mathbf{p})$. If the λ -levels of $G(\lambda)$ are scanned top-down, every new distinct connected component that appears corresponds to a local maxima at approximately density level λ . By then taking the corresponding connected components instead in $G((1-\beta)\lambda)$, with the $(1-\beta)$ multiplicative factor allowing for fluctuations in local density, we can identify cluster cores as the regions surrounding local modes.

In our pipeline, we take a top-down approach to tree segmentation. We first project the height-normalized cloud onto the XY-plane, similarly to Nelson et al. [28]. We



Fig. 3: An image of one of the forest plots in Switzerland. April tags are placed on individual trees to allow easy association of the measured traits.

then identify the core sets, reasoning that the trunks should primarily correspond to the densest regions. These cluster core sets are unprojected back to their corresponding 3D points. Subsequently, the remaining points are clustered via the second step of Quickshift++ wherein each point is moved closer to its nearest neighbor with the highest density $f_k(\mathbf{p})$ until it reaches some cluster core. The two main parameters of the approach are k for the k -th nearest neighbor and β for the fluctuation in density allowed in a cluster.

C. Geometric Primitive Fitting

For each point to be assigned to an instance, we filter out the points with a height between 1 m and 4 m. We then fit a cylinder following a RANSAC scheme, where we first sample five points and calculate the corresponding mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The mean defines the center of the cylinder, the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{\Sigma}$ defines its axis \mathbf{a} , and the square root of the second-largest eigenvalue defines the radius r . We define the error e_p per point as

$$e_p = \|\mathbf{a} \times (\mathbf{p} - \boldsymbol{\mu})\|_2 - r. \quad (6)$$

We take as inliers the points for which the error is less than 10% of the voxel size. Using this estimate as an initial guess, we then refine it through a least-squares minimization scheme. The DBH of a tree is then simply the diameter of the estimated cylinder. While some approaches [13], [30] first filter out points close to the nominal DBH height (1.3 m) and then fit a circle to the points, we find from our experiments (Sec. IV-C) a strong agreement with reference measurements with this alternative approach as well.

IV. EXPERIMENTAL EVALUATION

The main focus of this work is a pipeline to support automated forest inventory using mobile robotic laser scanning platforms. We present here our experiments showing the capabilities of our method to segment trees and compute the DBH for every tree. The results of our experiments as presented below also support our key claims, which are: (i)

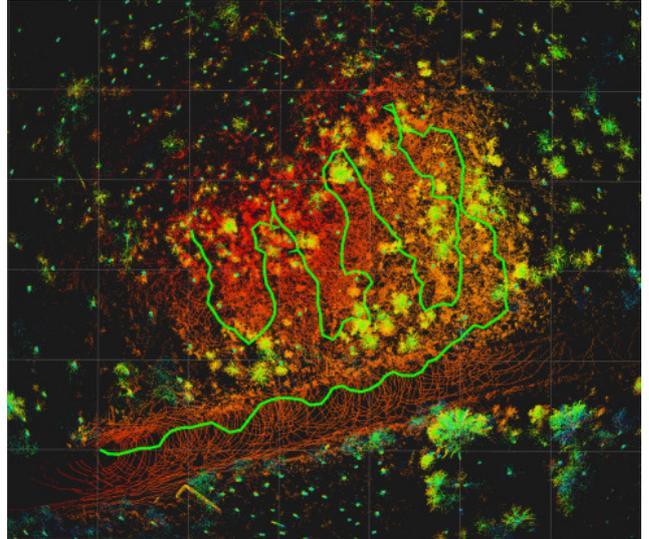


Fig. 4: The trajectory followed by the ANYmal robot during data collection in the forest near Evo, Finland. The platform was equipped with a Velodyne VLP-16.

strong performance in tree instance segmentation on mobile laser scans of forests, (ii) strong estimation of tree DBH through cylinder fitting, achieving high accuracy for forest inventory, and (iii) solid qualitative performance across data from varied sensor types with simple tuning.

A. Experimental Setup

We carried out field campaigns and collected extensive data from forests near Stein am Rhein, Switzerland, and Evo, Finland during March and May 2023 respectively under the supervision of the Swiss Federal Institute for Forest, Snow, and Landscape Research. LiDAR sensor rigs included a Hesai Pandar XT-32, Leica's BLK sensors, and the ANYmal robotic platform mounted with a Velodyne VLP-16. We show qualitative results on data from across these sensors and forests and focus our quantitative evaluation on the data collected from the Hesai sensor during the Switzerland campaign. Before mapping or measuring the trees, we attached April tags to the barks of the trees under study, noting the tag ID alongside any tree traits measured. This allowed a simplified association of the traits to laser scans by detecting the tags from RGB cameras attached to the Hesai sensor, giving a pose estimate of each tree in the map frame. With our approach, we aimed to keep the need for manual labeling of point cloud data, which is a tedious task, to as minimal as possible. However, we did label point clouds from 5 forest plots consisting of 143 individual tree instances, focusing on labeling only the trees for which ground truth traits were measured. Of this, we take a single forest plot with 13 labeled trees to optimize the parameters of our approach (two parameters) and those of the baselines when feasible. This plot is not included in testing, and we report the results in the following sections from experiments on the other four forest plots. We note that such an optimization step is completely optional in normal usage, and necessary only for fine-tuning.

Method	Plot 1	Plot 2	Plot 3	Plot 4	Average
Silva et al. [36]	0.197	0.355	0.163	0.384	0.275
Dalponte et al. [5]	0.194	0.365	0.159	0.400	0.280
Li et al. [18]	0.390	0.485	0.341	0.535	0.437
Donager et al. [7]	0.612	0.561	0.330	0.485	0.497
Ours	0.761	0.722	0.489	0.781	0.688

TABLE I: Evaluation of instance segmentation of different approaches comparing panoptic quality (PQ).

B. Instance Segmentation

The first experiment evaluates the performance of our instance segmentation approach. Its outcome supports the claim that we can accurately delineate trees along with their canopies into individual instances. We show comparisons against the methods of Silva et al. [36], Dalponte et al. [5], Li et al. [18], and Donager et al. [7]. The motivation for these baselines is that, in part, they are state-of-the-art methods, but also that their working implementation is available open-source. We do not compare against the results of Krisanski et al. [16] since their method does not perform canopy separation. For the forest plots under study, DBH measurements were collected only for a subset of trees in each plot. We labeled only this subset of trees, resulting in a dataset containing partially annotated point clouds. To then evaluate the segmentation performance of the approaches, we use the standard panoptic quality (PQ) metric [4], [15]. PQ is given by

$$PQ = \frac{\sum_{(p,g) \in TP} \text{IoU}(p,g)}{|\text{TP}| + \frac{1}{2}|\text{FP}| + \frac{1}{2}|\text{FN}|} \quad (7)$$

where p, g represent the predicted and ground truth instance labels, TP is the set of true positives, FP is the set of false positives, FN is the set of false negatives, IoU is the intersection over union, and $|S|$ represents the cardinality of the set S . This metric measures how well are the points assigned to their instances allowing us to evaluate the methods with respect to the partial ground truth annotations.

Tab. I shows a quantitative comparison of our approach against the baselines. We see that our approach outperforms the baselines by a substantial margin. We show further qualitative results of our approach in Sec. IV-D.

C. Tree Trait Estimation

The second experiment evaluates the tree trait estimation and illustrates that our approach can estimate DBH to produce an accurate forest inventory. We take the methods of Donager et al. [7] and Krisanski et al. [16] as baselines for DBH estimation and report the root-mean-square error (RMSE) values.

Tab. II presents a quantitative comparison of our approach against the baselines. As we can see from the results, our approach achieves around 11.8 cm RMSE in the DBH estimation. While we do not perform on-par with the approach of Krisanski et al. [16], their method exploits semantic segmentation to include only the tree’s stem into each instance, excluding any canopy or undergrowth vegetation.

Method	Plot 1	Plot 2	Plot 3	Plot 4	Average
Donager et al. [7]	2.988	2.721	2.963	-	2.890
Krisanski et al. [16]	0.031	0.051	0.100	0.021	0.051
Ours	0.088	0.116	0.182	0.054	0.110

TABLE II: Comparison of DBH estimation performance in RMSE (m) of different algorithms.

Our method, however, segments also the tree’s canopy along with the stem into each instance. Moreover, since we do not have semantic information, we potentially segment the bushes and undergrowth vegetation along with the canopy. To mitigate this, we clip the height-normalized point cloud below 1 m as a pre-processing step to remove such vegetation but we still found it insufficient in the forest plots that we studied. Simply clipping more of the cloud also renders the DBH estimation inefficient, as this is typically measured at around 1.3 m. As a consequence, our cylinder fitting strategy is not restricted to the stem, which degrades the performance in terms of the DBH estimation. Notice, however, that the gap in estimation accuracy is, on average, below 6 cm. The other geometric method proposed by Donager et al. [7] fails on this data with an average RMSE of 2.89 m. We exclude the result for Plot 4 from the average and in Tab. II since, in this case, Donager et al. [7] estimate an average DBH of over 300 meters. This is due to many outliers in their tree instance segmentation. We notice that, in general, the rather low performance of Donager et al. [7] stems from their instance segmentation method overestimating the number of trees in each forest plot, highlighting the importance of instance segmentation in forestry traits estimation.

D. Qualitative Performance of Our Approach

We show in Fig. 5 the result of our approach to cylinder fitting on a forest plot in Evo, Finland. Furthermore, in Fig. 6, we present the qualitative analysis of our pipeline showcasing the robustness of our approach. Note that the data for these experiments was collected from several sensors and from different geographical locations, i.e., forests in Switzerland and Finland.

In summary, our evaluation suggests that our method is applicable for producing forest inventory from data collected with mobile robotic platforms, providing accurate segmentation of trees and DBH estimation. The method is robust and requires minimal effort to be adapted to a different sensor or geographical region. Thus, we supported all our claims with this experimental evaluation.

V. CONCLUSION

In this paper, we presented a novel approach to instance segmentation of forest scenes mapped using mobile laser scanning platforms and subsequent estimation of per-tree traits. Our approach has only a handful of parameters, making it simple to tune and adapt. We implemented and quantitatively evaluated our approach on data collected using a mobile robotics sensor rig from a forest in Switzerland. Furthermore, we showcase good qualitative performance on



Fig. 5: Qualitative results of our cylinder fitting approach. Cylinders are shown in red. The point cloud has been cropped along its height to allow clearer visualization.

data collected in an entirely different type of forest in Finland and also on data from varied sensor setups. We provided comparisons to other existing techniques and supported all claims made in this paper. The experiments suggest that our approach has the potential to vastly speed up the generation of accurate inventory of wide expanses of forests, taking advantage of the cost-effectiveness and flexibility of mobile robotic platforms. Our approach outperforms popular existing baseline methods for tree instance segmentation. In future work, we aim to introduce semantics in our pipeline to distinguish between canopy, stem and undergrowth vegetation and further improve the DBH estimation.

ACKNOWLEDGMENTS

We thank Saurabh Gupta for proofreading the manuscript and providing visualizations of the results. We thank Leica Geosystems AG for supporting us with the BLK sensors. We thank also PreFor Oy and especially Henri Riihimäki for their assistance during the Finland data collection campaign. We thank Christian Ginzler and his remote sensing team from WSL for supporting the data collection in Switzerland by providing the drone DJI M600 and the TLS Rigl VZ-400i.

REFERENCES

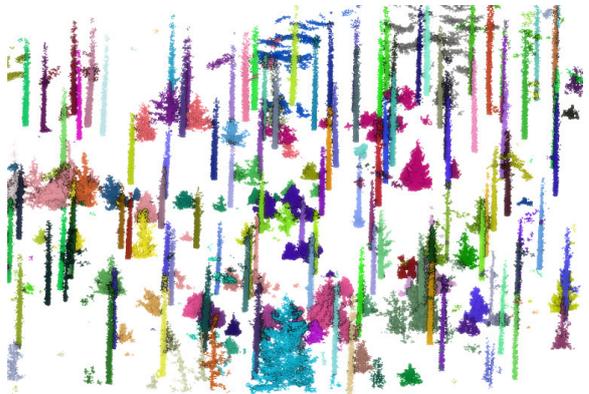
- [1] A. Burt, M. Disney, and K. Calders. Extracting individual trees from lidar point clouds using treeseg. *Methods in Ecology and Evolution*, 10(3):438–445, 2019.
- [2] N. Chebrolu, T. Läbe, O. Vysotska, J. Behley, and C. Stachniss. Adaptive Robust Kernels for Non-Linear Least Squares Problems. *IEEE Robotics and Automation Letters (RA-L)*, 6:2240–2247, 2021.
- [3] S.W. Chen, G.V. Nardari, E.S. Lee, C. Qu, X. Liu, R.A.F. Romero, and V. Kumar. Sloam: Semantic lidar odometry and mapping for forest inventory. *IEEE Robotics and Automation Letters (RA-L)*, 5(2):612–619, 2020.
- [4] B. Cheng, I. Misra, A.G. Schwing, A. Kirillov, and R. Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [5] M. Dalponte and D.A. Coomes. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods in Ecology and Evolution*, 7(10):1236–1245, 2016.



(a)



(b)



(c)

Fig. 6: Qualitative results of our tree instance segmentation pipeline. (a) shows the instance segmentation for the Leica’s BLK sensor on the Evo site in Finland, (b) for the Hesai LiDAR in the Switzerland site, and (c) for the ANYmal platform in the Evo site.

- [6] J.G. de Tanago, A. Lau, H. Bartholomeus, M. Herold, V. Avitabile, P. Raunonen, C. Martius, R.C. Goodman, M. Disney, S. Manuri, A. Burt, and K. Calders. Estimation of above-ground biomass of large tropical trees with terrestrial lidar. *Methods in Ecology and Evolution*, 9(2):223–234, 2018.
- [7] J.J. Donager, A.J.S. Meador, and R.C. Blackburn. Adjudicating perspectives on forest structure: How do airborne, terrestrial, and mobile lidar-derived estimates compare? *Remote Sensing*, 13(12), 2021.
- [8] S. Du, R. Lindenbergh, H. Ledoux, J. Stoter, and L. Nan. Adtree: Accurate, detailed, and automatic modelling of laser-scanned trees.

- Remote Sensing*, 11(18), 2019.
- [9] J.M. Fortin, O. Gamache, V. Grondin, F. Pomerleau, and P. Giguère. Instance segmentation for autonomous log grasping in forestry operations. In *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2022.
 - [10] P.D. Frenne, J. Lenoir, M. Luoto, B.R. Scheffers, F. Zellweger, J. Aalto, M.B. Ashcroft, D.M. Christiansen, G. Decocq, K.D. Pauw, S. Govaert, C. Greiser, E. Gril, A. Hampe, T. Jucker, D.H. Klings, I.A. Koelemeijer, J.J. Lembrechts, R. Marrec, C. Meeussen, J. Ogée, V. Tyystjärvi, P. Vangansbeke, and K. Hylander. Forest microclimates and climate change: Importance, drivers and future research agenda. *Global Change Biology*, 27(11):2279–2297, 2021.
 - [11] G. Grisetti, R. Kümmerle, C. Stachniss, and W. Burgard. A tutorial on graph-based SLAM. *IEEE Trans. on Intelligent Transportation Systems Magazine*, 2:31–43, 2010.
 - [12] S. Gupta, T. Guadagnino, B. Mersch, I. Vizzo, and C. Stachniss. Effectively Detecting Loop Closures using Point Cloud Density Maps. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2024.
 - [13] H.K. Heo, D.K. Lee, J.H. Park, and J.H. Thorne. Estimating the heights and diameters at breast height of trees in an urban park and along a street using mobile lidar. *Landscape and Ecological Engineering*, 15(3):253–263, 2019.
 - [14] H. Jiang, J. Jang, and S. Kpotufe. Quickshift++: Provably good initializations for sample-based mean shift. In *Proc. of the Intl. Conf. on Machine Learning (ICML)*, 2018.
 - [15] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár. Panoptic Segmentation. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.
 - [16] S. Krisanski, M.S. Taskhiri, S.G. Aracil, D. Herries, A. Muneri, M.B. Gurung, J. Montgomery, and P. Turner. Forest structural complexity tool—an open source, fully-automated tool for measuring forest point clouds. *Remote Sensing*, 13(22), 2021.
 - [17] S. Krisanski, M.S. Taskhiri, S.G. Aracil, D. Herries, and P. Turner. Sensor agnostic semantic segmentation of structurally diverse and complex forest point clouds using deep learning. *Remote Sensing*, 13(8), 2021.
 - [18] W. Li, Q. Guo, M.K. Jakubowski, and M. Kelly. A new method for segmenting individual trees from the lidar point cloud. *Photogrammetric Engineering and Remote Sensing (PE&RS)*, 78(1):75–84, 2012.
 - [19] X. Liang, J. Hyyppä, H. Kaartinen, M. Lehtomäki, J. Pyörälä, N. Pfeifer, M. Holopainen, G. Broly, P. Francesco, J. Hackenberg, H. Huang, H.W. Jo, M. Katoh, L. Liu, M. Mokroš, J. Morel, K. Olofsson, J. Poveda-Lopez, J. Trochta, D. Wang, J. Wang, Z. Xi, B. Yang, G. Zheng, V. Kankare, V. Luoma, X. Yu, L. Chen, M. Vastaranta, N. Saarinen, and Y. Wang. International benchmarking of terrestrial laser scanning approaches for forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*, 144:137–179, 2018.
 - [20] X. Liang, V. Kankare, J. Hyyppä, Y. Wang, A. Kukko, H. Haggren, X. Yu, H. Kaartinen, A. Jaakkola, F. Guan, M. Holopainen, and M. Vastaranta. Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*, 115:63–77, 2016.
 - [21] M. Liu, Z. Feng, Z. Zhang, C. Ma, M. Wang, B. ling Lian, R. Sun, and L. Zhang. Development and evaluation of height diameter at breast models for native chinese metasequoia. *PLOS ONE*, 12(8):1–16, 2017.
 - [22] X. Liu, G.V. Nardari, F.C. Ojeda, Y. Tao, A. Zhou, T. Donnelly, C. Qu, S.W. Chen, R.A.F. Romero, C.J. Taylor, and V. Kumar. Large-scale autonomous flight with real-time semantic slam under dense forest canopy. *IEEE Robotics and Automation Letters*, 7(2):5512–5519, 2022.
 - [23] J.A. Lutz, T.J. Furniss, D.J. Johnson, S.J. Davies, D. Allen, A. Alonso, K.J. Anderson-Teixeira, A. Andrade, J. Baltzer, K.M.L. Becker, E.M. Blomdahl, N.A. Bourg, S. Bunyavejchewin, D.F.R.P. Burslem, C.A. Cansler, K. Cao, M. Cao, D. Cárdenas, L.W. Chang, K.J. Chao, W.C. Chao, J.M. Chiang, C. Chu, G.B. Chuyong, K. Clay, R. Condit, S. Cordell, H.S. Dattaraja, A. Duque, C.E.N. Ewango, G.A. Fischer, C. Fletcher, J.A. Freund, C. Gardina, S.J. Germain, G.S. Gilbert, Z. Hao, T. Hart, B.C.H. Hau, F. He, A. Hector, R.W. Howe, C.F. Hsieh, Y.H. Hu, S.P. Hubbell, F.M. Inman-Narahari, A. Itoh, D. Janík, A.R. Kassim, D. Kenfack, L. Korte, K. Král, A.J. Larson, Y. Li, Y. Lin, S. Liu, S. Lum, K. Ma, J.R. Makana, Y. Malhi, S.M. McMahon, W.J. McShea, H.R. Memiaghe, X. Mi, M. Morecroft, P.M. Musili, J.A. Myers, V. Novotny, A. de Oliveira, P. Ong, D.A. Orwig, R. Ostertag, G.G. Parker, R. Patankar, R.P. Phillips, G. Reynolds, L. Sack, G.Z.M. Song, S.H. Su, R. Sukumar, I.F. Sun, H.S. Suresh, M.E. Swanson, S. Tan, D.W. Thomas, J. Thompson, M. Uriarte, R. Valencia, A. Vicentini, T. Vrška, X. Wang, G.D. Weiblen, A. Wolf, S.H. Wu, H. Xu, T. Yamakura, S. Yap, and J.K. Zimmerman. Global importance of large-diameter trees. *Global Ecology and Biogeography*, 27(7):849–864, 2018.
 - [24] C. Malzer and M. Baum. A hybrid approach to hierarchical density-based cluster selection. In *Intl. Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2020.
 - [25] E. Marks, M. Sodano, F. Magistri, L. Wiesmann, D. Desai, R. Marcuzzi, J. Behley, and C. Stachniss. High precision leaf instance segmentation for phenotyping in point clouds obtained under real field conditions. *IEEE Robotics and Automation Letters (RA-L)*, 8(8):4791–4798, 2023.
 - [26] S.M.K. Moorthy, K. Calders, M.B. Vicari, and H. Verbeeck. Improved supervised learning-based approach for leaf and wood classification from lidar point clouds of forests. *IEEE Trans. on Geoscience and Remote Sensing*, 58(5):3057–3070, 2020.
 - [27] A.S. Mori, K.P. Lertzman, and L. Gustafsson. Biodiversity and ecosystem services in forest ecosystems: a research agenda for applied forest ecology. *Journal of Applied Ecology*, 54(1):12–27, 2017.
 - [28] H.J. Nelson and N. Papanikolopoulos. Pre-clustering point clouds of crop fields using scalable methods. *arXiv preprint*, arXiv:2107.10950, 2022.
 - [29] L. Pellissier, M. Anzini, L. Maiorano, A. Dubuis, J. Pottier, P. Vittoz, and A. Guisan. Spatial predictions of land-use transitions and associated threats to biodiversity: the case of forest regrowth in mountain grasslands. *Applied Vegetation Science*, 16(2):227–236, 2013.
 - [30] A. Proudman, M. Ramezani, S. Digumarti, N. Chebrolu, and M. Fallon. Towards real-time forest inventory using handheld LiDAR. *Journal on Robotics and Autonomous Systems (RAS)*, 157, 2022.
 - [31] C. Qi, K. Yi, H. Su, and L.J. Guibas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In *Proc. of the Conf. on Neural Information Processing Systems (NeurIPS)*, 2017.
 - [32] M. Ramezani, M. Mattamala, and M. Fallon. AEROS: Adaptive Robust Least-Squares for Graph-Based SLAM. *Frontiers in Robotics and AI*, 2022.
 - [33] G. Roggiolani, M. Sodano, T. Guadagnino, F. Magistri, J. Behley, and C. Stachniss. Hierarchical approach for joint semantic, plant instance, and leaf instance segmentation in the agricultural domain. In *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2023.
 - [34] J.R. Roussel, D. Auty, N.C. Coops, P. Tompalski, T.R. Goodbody, A.S. Meador, J.F. Bourdon, F. de Boissieu, and A. Achim. lidar: An r package for analysis of airborne laser scanning (als) data. *Remote Sensing of Environment*, 251:112061, 2020.
 - [35] F. Schiefer, T. Kattenborn, A. Frick, J. Frey, P. Schall, B. Koch, and S. Schmidlein. Mapping forest tree species in high resolution uav-based rgb-imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*, 170:205–215, 2020.
 - [36] C.A. Silva, A.T. Hudak, L.A. Vierling, E.L. Loudermilk, J.J. O’Brien, J.K. Hiers, S.B. Jack, C. Gonzalez-Benecke, H. Lee, M.J. Falkowski, and A. Khosravipour. Imputation of individual longleaf pine (*Pinus palustris* mill.) tree attributes from field and lidar data. *Canadian Journal of Remote Sensing*, 42(5):554–573, 2016.
 - [37] C. Stachniss, J. Leonard, and S. Thrun. *Springer Handbook of Robotics*, 2nd edition, chapter Chapt. 46: Simultaneous Localization and Mapping. Springer Verlag, 2016.
 - [38] C. Sun, C. Huang, H. Zhang, B. Chen, F. An, L. Wang, and T. Yun. Individual tree crown segmentation and crown width extraction from a heightmap derived from aerial laser scanning data using a deep learning framework. *Frontiers in Plant Science*, 13, 2022.
 - [39] A. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. In *Proc. of the Europ. Conf. on Computer Vision (ECCV)*, 2008.
 - [40] I. Vizzo, T. Guadagnino, B. Mersch, L. Wiesmann, J. Behley, and C. Stachniss. KISS-ICP: In Defense of Point-to-Point ICP – Simple, Accurate, and Robust Registration If Done the Right Way. *IEEE Robotics and Automation Letters (RA-L)*, 8(2):1029–1036, 2023.
 - [41] O. Vysotska and C. Stachniss. Lazy Data Association For Image Sequences Matching Under Substantial Appearance Changes. *IEEE Robotics and Automation Letters (RA-L)*, 1(1):213–220, 2016.
 - [42] D. Wang, S.M. Takoudjou, and E. Casella. Lewos: A universal leaf-wood classification method to facilitate the 3d modelling of large

tropical trees using terrestrial lidar. *Methods in Ecology and Evolution*, 11(3):376–389, 2020.

- [43] H. Weiser, J. Schäfer, L. Winiwarter, N. Krašovec, F.E. Fassnacht, and B. Höfle. Individual tree point clouds and tree measurements from multi-platform laser scanning in german forests. *Earth System Science Data*, 14(7):2989–3012, 2022.
- [44] M.A. Wulder, J.C. White, R.F. Nelson, E. Næsset, H.O. Ørka, N.C. Coops, T. Hilker, C.W. Bater, and T. Gobakken. Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment*, 121:196–209, 2012.
- [45] W. Zhang, J. Qi, P. Wan, H. Wang, D. Xie, X. Wang, and G. Yan. An easy-to-use airborne lidar data filtering method based on cloth simulation. *Remote Sensing*, 8(6), 2016.