

Article

NeRF-Accelerated Ecological Monitoring in Mixed-Evergreen Redwood Forest

Adam Korycki ^{1,*} , Cory Yeaton ², Gregory S. Gilbert ³ , Colleen Josephson ¹ , Steve McGuire ¹

¹ Electrical and Computer Engineering Department, University of California Santa Cruz; 1156 High Street, Santa Cruz, CA 95064, USA; cjosephson@ucsc.edu (C.J.); stmccguir@ucsc.edu (S.M.)

² Department of Ecology and Evolutionary Biology, University of California Santa Cruz; 1156 High Street, Santa Cruz, CA 95064, USA; cyeaton@ucsc.edu

³ Department of Environmental Studies, University of California Santa Cruz; 1156 High Street, Santa Cruz, CA 95064, USA; ggilbert@ucsc.edu (G.S.G.)

* Correspondence: akorycki@ucsc.edu

Abstract: Forest mapping provides critical observational data needed to understand the dynamics of forest environments. Notably, tree diameter at breast height (DBH) is a metric used to estimate forest biomass and carbon dioxide (CO₂) sequestration. Manual methods of forest mapping are labor intensive and time consuming, a bottleneck for large-scale mapping efforts. Automated mapping relies on acquiring dense forest reconstructions, typically in the form of point clouds. Terrestrial laser scanning (TLS) and mobile laser scanning (MLS) generate point clouds using expensive LiDAR sensing and have been used successfully to estimate tree diameter. Neural radiance fields (NeRFs) are an emergent technology enabling photorealistic, vision-based reconstruction by training a neural network on a sparse set of input views. In this paper, we present a comparison of MLS and NeRF forest reconstructions for the purpose of trunk diameter estimation in a mixed-evergreen Redwood forest. In addition, we propose an improved DBH-estimation method using convex-hull modeling. Using this approach, we achieved 1.68 cm RMSE (2.81%), which consistently outperformed standard cylinder modeling approaches.

Keywords: 3D forest reconstruction; neural radiance fields (NeRF); LiDAR; Simultaneous Localization and Mapping (SLAM); diameter at breast height (DBH); tree modeling



Academic Editors: Zengxin Zhang and Guojie Wang

Received: 3 December 2024

Revised: 8 January 2025

Accepted: 10 January 2025

Published: 17 January 2025

Citation: Korycki, A.; Yeaton, C.; Gilbert G.; Josephson, C.; McGuire S. NeRF-Accelerated Ecological Monitoring in Mixed-Evergreen Redwood Forest. *Forests* **2025**, *16*, 173. <https://doi.org/10.3390/f16010173>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forests are the Earth's largest terrestrial carbon store, holding more than three decades worth of global CO₂ emissions [1] and consuming a quarter of new anthropogenic emissions [2]. Pressingly, climatic trends are revealing grave uncertainty for long-term stability. According to U.S. Forest Service aerial surveys, over 200 million trees died in California since 2010, with 62 million dead in 2016 alone [3]. The warming climate and the consequence of longer, more severe drought cycles are the primary culprits of this mass die-off. Significant numbers of dead and dying trees dramatically increase the risk of wildfires; these counts do not include tree deaths caused by wildfires, which adds hundreds of millions to the toll.

Forest management is a recognized, cost-effective approach to mitigating the effects of the climate crisis [4]. Global carbon accounting, which is a crucial contribution to informed climate change policies, relies on large-scale forest surveys. Tree diameter at breast height (DBH) is a primary data point used in ecological monitoring and carbon accounting efforts; the conventional means of determining DBH relies on a human forester with a measuring

tape. Accurate, automated methods of DBH estimation could drastically reduce the time and effort needed to perform surveys, opening doors for large-scale mapping efforts.

Three-dimensional (3D) reconstruction is the task of digitally representing real-world settings, typically in the form of a point cloud. Metric reconstruction is a subclass that preserves true scale in the recovered geometry and affords the ability to indirectly measure scene features (e.g., volume, length). Terrestrial laser scanning (TLS) is a well-studied approach for mapping forest inventories, offering the potential for rapid ecological assessment. However, these systems are expensive, costing 80,000–20,000 USD, and require a skilled operator. Survey-grade TLS systems can provide centimeter-level diameter estimation in forests with uniform tree structures and even terrain [5]. The main technical issue faced by TLS methods is tree occlusions that require stitching many scans together from different spatial locations, a critical step in recovering complex geometries of forests. Research toward using mobile robot platforms in combination with Simultaneous Localization and Mapping (SLAM) algorithms addresses this problem using optimized pose-graphs to align thousands or millions of LiDAR scans taken along a robot’s trajectory. A 2017 paper [6] using SLAM reports a best-case DBH estimation RMSE of 2.38 cm for well-represented trees. A 2024 study [7] achieved 1.93 cm RMSE using a mixed Hugh-RANSAC trunk modeling approach. However, these solutions require expensive 3D LiDAR and inertial measurement unit (IMU) hardware (10,000–25,000 USD).

Recent advances in the fields of computer vision and deep learning offer a new paradigm for generating 3D reconstructions. Neural Radiance Fields (NeRFs) [8] are an emergent technology enabling the recovery of complex 3D geometry by training a neural network on conventional imagery. NeRFs are a remarkable advance over traditional photogrammetry, producing higher quality, photorealistic 3D reconstructions from sparser input imagery and at an accessible efficiency. Since 2020, a community of developers has contributed hundreds of methodological improvements, rapidly improving its performance and accessibility. The ability to export NeRFs as point clouds lends itself as an aggressive alternative to expensive LiDAR-based mapping.

We present an evaluation of NeRF-based forest reconstruction for the task of DBH estimation of mixed-evergreen redwood forest located in Santa Cruz, California. This study compares NeRF reconstructions trained on conventional mobile phone imagery to LiDAR-inertial SLAM reconstructions sourced from a quadruped robot equipped with a custom multi-modal sensing platform. This paper also expands the viability and accuracy of TreeTool [9], a Python toolkit for rendering DBH estimates from forest point clouds. Specifically, we propose a new set of features to support robust tree detection and accurate DBH estimation. Many studies have relied on cylinder-fitting to model trunk morphology. We observe a DBH underestimation trend with this method and propose an improved convex-hull modeling approach. In summary, the contributions of this work are:

- Quantitative field study evaluating the performance of NeRF-based forest reconstructions compared to LiDAR-inertial SLAM with regards to DBH estimation accuracy.
- Improved DBH estimation accuracy via a trunk modeling approach using convex-hull and density-based filtering methods.
- Open-source modeling code and forest datasets, including SLAM and NeRF reconstructions of a mixed-evergreen Redwood forest, are freely available at <https://github.com/harelab-ucsc/RedwoodNeRF> (accessed on 7 January 2025).

2. Theoretical Background

2.1. The SLAM Approach

The SLAM problem can be broken into two tasks: building a map of the environment and simultaneously estimating the robot’s trajectory within that map. More specifically,

given all sensor measurements $z_{1:T}$, all robot motion commands $u_{1:T}$, and an initial robot position x_0 , estimate the posterior probability of the complete robot trajectory $x_{1:T}$ and a globally consistent map $\mathbf{m}$ of the explored environment [10]:

$$p(x_{1:T}, \mathbf{m} | z_{1:T}, u_{1:T}, x_0) \quad (1)$$

This Bayesian problem formulation benefits from the ability to fuse several sensing modalities together while accounting for individual sensor noise. The modern-day solution to SLAM formulates the problem as a factor graph [11] where each node represents a robot pose x_i (relative 3D spatial location and orientation) and each edge represents a transformation T_i between nodes. Edges also encapsulate loop-closure constraints. Loop-closure is the subprocess of identifying previously observed landmarks in order to correct for drift and error accumulation in the estimated trajectory. Based on the prior pose-graph, a global optimization is used to minimize the estimation error given the sensor measurements and loop-closure constraints. The resulting pose corrections are back-propagated, resulting in reduced error of scan alignment and dense scene reconstruction.

2.2. NeRF Scene Representation

NeRF [8] is a current state-of-the-art solution to the problem of novel view synthesis. This problem involves generating an image of a 3D scene from a particular view when the only available information is images from other views. Two pivotal ingredients to the NeRF method are continuous volumetric representation and deep, fully-connected network architecture. NeRFs inherit the exceptional photorealism and reconstruction fidelity of continuous representation at a fraction of the storage cost compared to discrete approaches using voxels or meshes [8]. The scene is modeled as a Multilayer Perceptron Network (MLP), which takes a 5D input vector composed of spatial location $\mathbf{X} = (x, y, z)$ and viewing angle $\mathbf{d} = (\theta, \phi)$ and learns the weights Θ to map each 5D coordinate to the corresponding 4D output vector of color $\mathbf{c} = (r, g, b)$ and volume density (σ). The 5D input space is sampled using ray tracing. The network architecture is two MLPs. The first learns only volume density based on input location. The second learns color based on location and viewing direction as well as σ . This enables multi-view consistency, critical for non-Lambertian lighting (lighting conditions that include high dynamic range across the scene). A positional encoding layer is used to better represent high-frequency color-density functions. Instead of performing volume rendering uniformly along the rays, the NeRF method uses hierarchical volume sampling to identify relevant regions of the scene and avoid excessive calculations in rendering free space. Figure 1 illustrates a high-level flow within the NeRF pipeline.

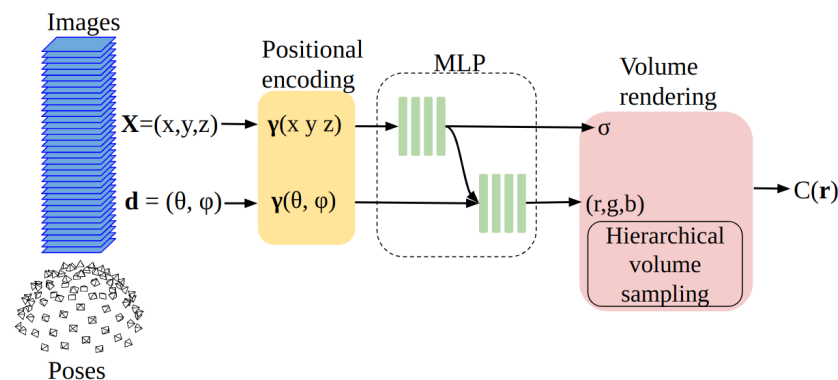


Figure 1. NeRF scene representation flow. Sparse images with corresponding poses are sampled using ray-tracing to generate 5D input vector comprised of location (x, y, z) and viewing direction (θ, ϕ) . A cascaded MLP learns the weights to map this 5D vector to output color (r, g, b) and volume density σ . Volume rendering composites the learned rays to novel views.

Scene reconstructions are rendered from the learned color-ray space by filtering low color-density regions to only leave surfaces. In order to extract and accurately measure geometric features, the image poses must be **metrically relevant** (in real-world distance units). Standard Structure from Motion (SfM) pipelines (such as COLMAP [12]) are not able to maintain real-world scale since absolute depth is not available from monocular visual odometry without additional information about the scene. Huang et al. [13] use SfM poses to generate NeRF reconstructions of two trees. To cope with scale ambiguity, the authors derive a scale factor from a ground truth TLS reconstruction, which is impractical in many real-world cases where this external information is unavailable. In this paper, we directly derive metrically accurate camera poses from visual-inertial (VI) SLAM, eliminating the need for any prior reconstruction.

3. Design and Methods

3.1. Mobile Laser Scanning via LiDAR-Inertial SLAM

In order to perform SLAM-based reconstruction, we designed a robot based on the Unitree B1 quadruped platform. Terrain maneuverability was a prioritized feature to cope with rough, uneven forest terrain and complex obstacles. A custom-built multi-modal sensor head is attached, which includes LiDAR, stereo vision, inertial, and GNSS+RTK sensing modes. For online processing, the robot is equipped with an external $\times 86$ mini computer (Commell LE-37R; Taiwan), which includes a 4.5 GHz Core i7-1270pe CPU, 64 GB RAM, and 2 TB storage. The LiDAR is an Ouster OS0-128 (Ouster, Inc; San Francisco, CA, USA) with a 90° vertical field-of-view and 128 horizontal channels. The IMU is an IMX5-RUG3 (Inertial Sense, Inc; Utah, USA) capable of 1 KHz output, and fused EKF attitude estimates.

The computer runs ROS2 Iron for sensor recording and running LiDAR inertial odometry smoothing and mapping (LIOSAM) [14]. This software fuses LiDAR and IMU data together to create dense spatial reconstructions in real-time along with optimized pose estimations. LIOSAM tightly couples LiDAR and inertial data in a joint optimization using a factor-graph SLAM architecture. Through loop-closure factors, LIOSAM is able to achieve minimal drift in large exploration volumes [14] (see Figure 2).

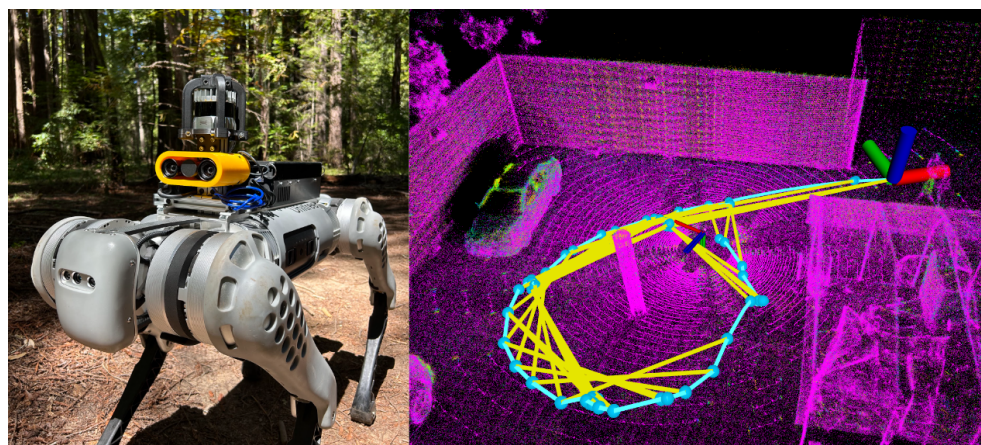


Figure 2. Quadruped robot creating a dense LiDAR-inertial reconstruction in a forest environment (left). LIOSAM visualization of estimated trajectory (torquoise), loop-closure events (yellow), and tightly aligned LiDAR scans (magenta) (right).

3.2. NeRF Reconstruction Pipeline

3.2.1. Training Data

Metrically relevant camera poses are needed in order to measure world features from NeRF reconstructions, while COLMAP can perform high quality vision-based reconstruc-

tion, it suffers scale ambiguity. The typical solution is to use VI or LiDAR-inertial SLAM pose estimation in which the metric information is derived from IMU or LiDAR range measurements. In this study, we use an iOS application called NeRFCapture [15] which uses Apple ARKit to provide camera poses in real time. ARKit uses VI odometry with multi-sensor fusion, which lends itself as a good option for metric pose estimation.

3.2.2. Software Implementation

The base NeRF method discussed in Section 2.2 has seen hundreds of proposed improvements over the years. Nerfacto [16] is a method that draws from several other methods [17–21] and has been shown to work well in a variety of in-the-wild settings. For this reason, we chose the Nerfacto method for this study.

Nerfacto improves on the base method in a few key dimensions, the first of which is **pose refinement**. Error in image poses results in cloudy artifacts and loss of sharpness in the reconstructed scene. The Nerfacto method uses the back-propagated loss gradients to optimize the poses for each training iteration. Another improvement is in the ray-sampling. Rays of light are modeled as conical frustums, and a **piece-wise sampling** step uniformly samples the rays up to a certain distance from the camera origin and then samples subsequent sections of the conical ray at step sizes that increase with each sample. This allows for high-detail sampling of close portions of the scene while efficiently sampling distant objects as well. The output is fed into a **proposal sampler**, which consolidates sample locations to sections of the scene that contribute most to the final 3D scene render. The output of these sampling stages is fed into the Nerfacto field, which incorporates **appearance embedding**, accounting for varying exposure among the training images.

We used the nerfstudio [22] API, which makes training and exporting NeRF reconstructions extremely simple. Posed image data is copied to a remote desktop PC for training. This computer hosts a 3.8 GHz AMD 3960X CPU (AMD; Taiwan), 64 GB RAM, and 2 TB storage. The PC is also outfitted with two NVIDIA RTX-3070 (Nvidia; Taiwan) graphics cards, which aggregate to 16 GB of VRAM. The system runs Ubuntu 22.04 with CUDA-11.8 to interface with GPU hardware.

3.3. Tree Segmentation and Modeling

3.3.1. TreeTool Framework

To process forest reconstructions and estimate tree DBH, we use TreeTool [9], a Python library built on Point Data Abstraction Library (PDAL) and Point Cloud Library Python (pclpy). TreeTool breaks the process down into three distinct steps covering filtering, detection, and modeling stages.

The goal of the filtering stage is to remove all non-trunk points, mainly the ground and foliage. The ground is segmented using an improved simple morphological filter (SMRF) proposed by Pingel et al. [23]. This technique uses image-inpainting to accurately model complex, uneven terrain. Once the ground points are removed, TreeTool uses a surface-normal filter to remove foliage points. This is based on their observation that the surfaces created by trunk points have horizontal normals. TreeTool also filters non-trunk points by considering the curvature of surfaces. Curvature is interpreted as the percentage of information held by the eigenvalue associated with the normal vector. High-curvature points are discarded, which removes foliage.

The detection stage groups the filtered trunk points into individual stem sections. TreeTool uses Euclidean-distance to perform nearest-neighbor clustering. Some points belonging to the same trunk are inevitably grouped into separate clusters due to occlusions and errors in the reconstruction. To cope with this, TreeTool groups same-trunk clusters together. We add an extra clustering step using density-based clustering applications with

noise (DBSCAN) [24], which addresses the case where points from different trunks are grouped together. This is especially prevalent for resprouting trees like redwoods, which commonly grow with conjoined trunk bases.

The last stage involves modeling the segmented and filtered trees to estimate diameter and location. Tree clusters are vertically cropped such that the remaining clusters represent the trunks between 1.0 and 1.6 m above the modeled ground surface. DBH is estimated by taking the maximum diameter reported between cylinder and ellipse fitting methods. Random sample consensus (RANSAC) fits a cylinder to the cropped trunk cluster. An additional ellipse model is generated using least-squares on a 2D projection of the points.

3.3.2. Convex Hull Modeling Approach

The use of RANSAC for modeling trees as cylinders to estimate DBH is common in the literature [6,7]. An advantage of this approach is the ability to extrapolate DBH from partially represented tree trunks, a common occurrence since optimal scene coverage is often not possible in complex forest terrain. These papers consider forests with uniform, cylindrical tree structure and an absence of near-ground trunk foliage, rendering their cylinder-approach as a viable modeling method with impressive accuracy. A downside of this method is that for well-represented trunks, a cylinder model is prone to underfitting the true trunk diameter. This is even more prevalent for tree species with deeply furrowed, irregular bark texture. Another limitation of this method is the inability to model irregular, bowed trunk shapes.

We propose a modeling approach that considers tree point clouds as stacks of convex-hull slices, as seen in Figure 3. We relax the morphological assumption of cylinder-modeling methods, which opens the possibility to model highly irregular trunk shapes. The trunk is vertically partitioned into 20 cm thick slices. Each slice is extracted and rotated to be collinear with the z-axis. By manipulating each slice independently, our method accounts for skewed, contorted trunk structures commonly found in non-coniferous forests. A 2D xy projection of the points is used to fit a convex-hull around the surface of the cloud, which emulates manual DBH measurement via girth tape. To deal with noisy points, we introduce a layer of DBSCAN that removes low-density regions. DBH is estimated by considering the slice at 1.3 m above the ground.

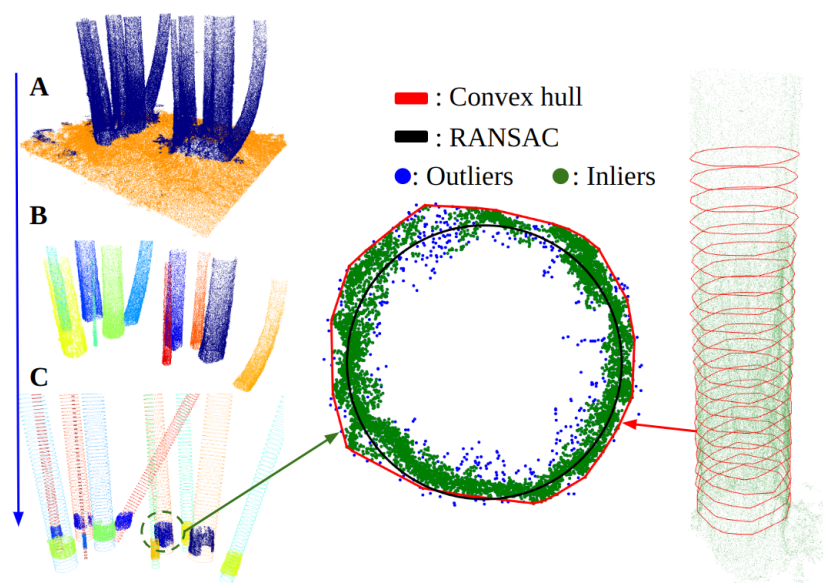


Figure 3. TreeTool process applied to a forest NeRF reconstruction. (A) ground segmentation, (B) trunk segmentation, and (C) trunk modeling. Our tree modeling approach considers trees as stacks of convex-hull slices which outperformed other approaches by 3–4× in terms of DBH estimation accuracy.

We take the maximum value across LS, RANSAC, and convex-hull methods as the final diameter to account for partial trunk cases. DBSCAN parameters control the maximum distance ϵ of points to be considered in a neighborhood, and the minimum point count (minPts) within that region to be considered a dense region. We found an ϵ range between 1 and 3 cm to have good outlier rejection on 2D trunk clouds. The minPts parameter is dependent on the 2D surface density $\rho_c [\frac{\text{points}}{\text{m}^2}]$ of the trunk cloud; we observed successful filtering in the range of 5–40 points.

3.4. Study Area and Data Collection

To validate the precision and accuracy of the proposed NeRF-derived convex-hull DBH method, we conducted an experiment in the Forest Ecology Research Plot (FERP) [25], a globally recognized ForestGEO [2] site in the Santa Cruz mountains along the central coast of California, USA. This plot spans 16 ha with over 51,000 recorded stem locations and DBH measurements. The forest census is repeated on a 5-year cycle.

The FERP is partitioned into 400 20×20 m subplots denoted by Ex_Ny, where x and y are the distance in meters from the SW corner of the FERP (37.012416, -122.074833) to the SW corner of the subplot. This study considered sections of forest in subplots E340_N360 and E340_N380. The data collection effort was accomplished over two visits and spans two datasets. LiDAR and IMU data were recorded at 10 Hz and 500 Hz, respectively. NeRF training data was collected using an iPhone 14 camera (1920 × 1440) and NeRFCapture [15]. Training via nerfstudio [22] lasted 300K iterations and took 15 min for both datasets. As a reference technique, DBH was taken manually via girth tape by a trained research assistant. The dataset parameters, including field-work duration across methods, are listed in Table 1.

3.4.1. Dataset A

In the first dataset, the robot was teleoperated around a cluster of 11 coast redwood trees (*Sequoia sempervirens*) to generate a SLAM reconstruction. The robot was also navigated through an opening between the trees to recover additional occluded trunk and ground geometry within the cluster. NeRF training imagery was collected by an untrained human,

traversing the tree cluster in a similar fashion. The trees had 360° coverage in the training data since nearby terrain afforded easier maneuvering.

3.4.2. Dataset B

The second dataset is larger by area, spanning the entire E340_N360 subplot which included more challenging terrain and foliage occlusions. This area consisted of 6 coast redwood and 3 Douglas-fir (*Pseudotsuga menziesii*) trees. The robot's obstacle-avoidance mode enabled maneuvering in complex terrain, but at a significantly reduced pace (Table 1). Stems less than 8 cm in diameter were not considered in this study, as robust DBH estimation was unstable in this size range.

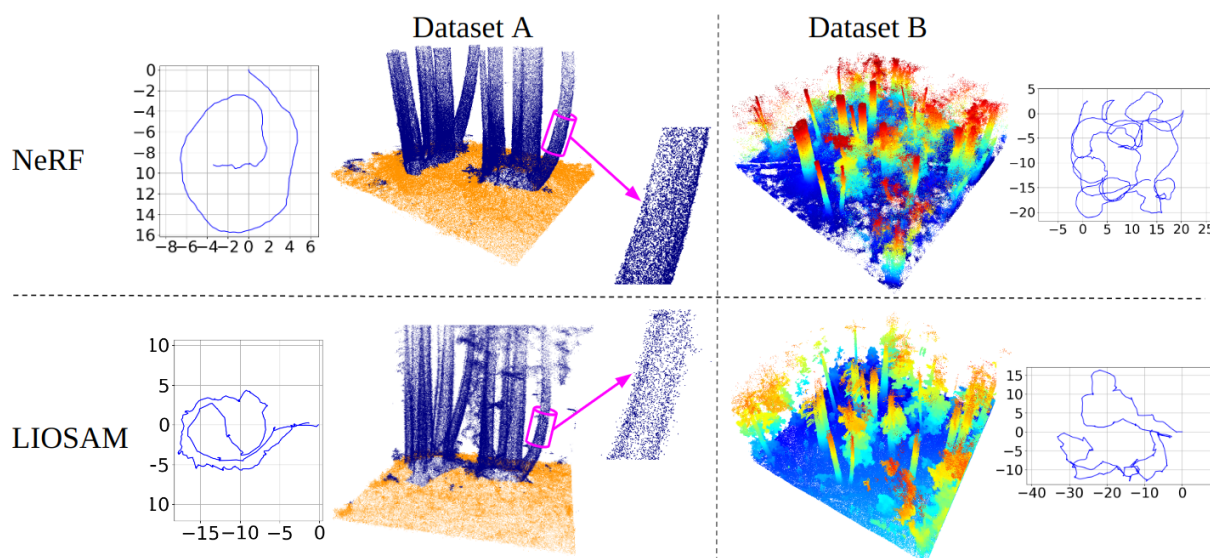


Figure 4. Forest reconstructions produced by SLAM (**bottom** row) and NeRF (**top** row) methods of both datasets. Adjacent plots are data collection trajectories for each reconstruction. In dataset A, we illustrate the effectiveness of segmentation between the ground points (orange) and trees (violet). We use a z-axis color gradient to enhance the visualization of dataset B reconstructions, as this region included more complex ground-level vegetation. The figure also compares a zoomed-in section of a tree trunk. The NeRF reconstruction is approximately 4× denser compared to SLAM, and is of higher surface quality.

Table 1. Study area parameters, quantity of recorded data, and a comparison of reconstruction density and fieldwork duration between NeRF and SLAM approaches. Duration includes field-validation of recorded data, not post processing.

Dataset	Area (m ²)	Tree Count	LiDAR Frames	Images	Fieldwork Duration			Point Count	
					NeRF	SLAM	Manual	NeRF	SLAM
A	140	11	2172	166	5 min	30 min	45 min	2.69 M	704 K
B	400	9	7498	847	8 min	40 min	52 min	26.87 M	7.83 M

4. Results

We assess the accuracy of DBH estimation for each dataset independently and combined, using bias, which gives an idea of over/under estimation-trends), root mean squared

error (RMSE), and standard deviation, as these are commonly referenced metrics in this domain. These are defined as:

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (y_i - y_{ri}) \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum (y_i - y_{ri})^2}{n}} \quad (3)$$

where y_i and y_{ri} are the estimated and reference diameters across n estimations. We also provide relative RMSE which is obtained by dividing the RMSE from (3) by the mean of the reference diameters.

The low bias values achieved with the NeRF convex-hull approach (−0.28 cm in Dataset A and −0.86 cm in Dataset B; Table 2) indicate minimal systematic error, compared to the significant under-estimation trends observed in RANSAC/LS (−4.35 cm and −4.59 cm, respectively; Table 2). Furthermore, the RMSE results, with a best-case value of 1.26 cm and an average of 1.68 cm across datasets, align well with the low bias values observed, demonstrating the reliability and consistency of our method.

Table 2. DBH estimation accuracy between our proposed convex-hull tree modeling approach and RANSAC for NeRF and SLAM reconstructions.

Metric	Dataset A				Dataset B			
	NeRF		SLAM		NeRF		SLAM	
	Convex-Hull	RANSAC	Convex-Hull	RANSAC	Convex-Hull	RANSAC	Convex-Hull	RANSAC
Bias (cm)	−0.28	−4.35	−1.35	−6.89	−0.86	−4.59	1.56	−3.69
RMSE (cm)	1.26	4.96	2.32	7.12	2.09	5.28	1.93	4.53
RMSE (%)	2.32	9.14	4.27	13.12	3.13	7.93	2.90	6.80
Std (cm)	1.29	2.49	1.97	1.86	2.02	2.77	0.93	2.88

The superior performance of the NeRF convex-hull method can be attributed to its ability to accurately model complex and irregular trunk geometries, which are prevalent in natural forests. In contrast, RANSAC’s reliance on predefined geometric primitives, such as cylinders and ellipses, leads to significant underfitting, particularly for trees with non-uniform diameters or irregular bark structures (Figure 5).

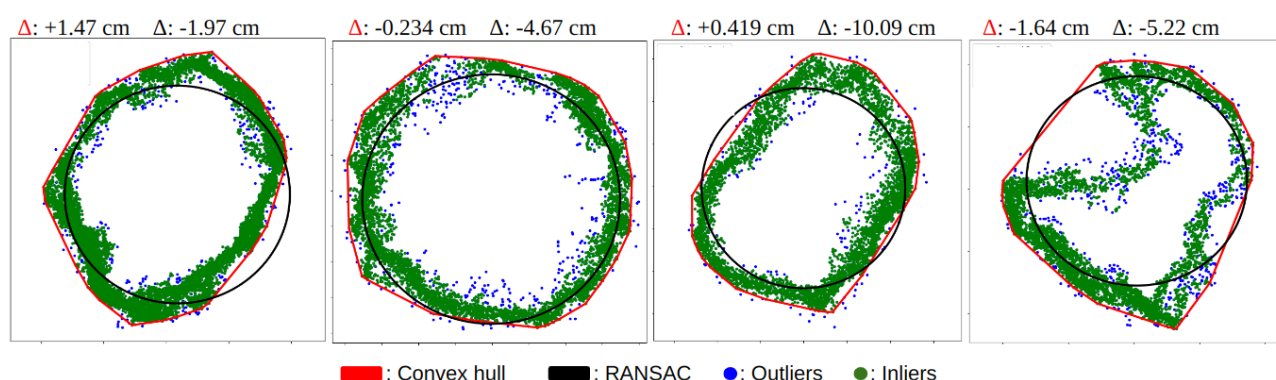


Figure 5. Four comparisons of RANSAC and convex-hull modeling approaches. Deltas between manual DBH and each modeling approach are provided on the top line. RANSAC cylinder modeling consistently under-fits well-represented trunk projections. Convex hull DBH estimation outperformed RANSAC by 3–4×.

Dataset B presented additional challenges due to complex terrain and darker light conditions, which could have impacted the quality of the NeRF reconstruction. Despite these challenges, the convex-hull method maintained robust performance, with only a marginal increase in RMSE from 1.26 cm in Dataset A to 2.09 cm in Dataset B.

The NeRF reconstructions were consistently $3\text{--}4\times$ more dense than the SLAM reconstructions (Table 1). The sparsity associated with LIOSAM is due to the nature of points being registered by laser pulse returns that have a resolution of 262 k points (128×2048) per LiDAR frame, a physical limitation of the hardware. NeRFs are capable of higher-density reconstruction since the geometry is rendered by sampling the learned color-ray space and filtering out low-density sections to only represent surfaces. This increased point density directly translates to finer surface representations and improved DBH estimation accuracy. This advantage is particularly pronounced in forests with intricate geometries, where the sparse point clouds generated by LIOSAM often fail to capture key structural details. By leveraging the ability of NeRFs to segment and filter high-density regions, our approach overcomes these limitations, resulting in more precise trunk modeling and DBH estimation.

5. Discussion

In comparison to the existing body of literature on automated DBH estimation, several studies provide a useful starting point for benchmarking and evaluation. Liang et al. [26] report a RMSE of 0.82 cm (4.21%) for automated DBH measurements compared to manual measurements obtained from TLS data. However, their analysis does not include a comparison of the automated estimates with ground truth stem measurements obtained using girth tape. Pierzchała et al. [6] achieve a RMSE of 2.38 cm (9%) compared to manual ground truth measurements using a mobile robot and LiDAR-inertial SLAM. Freißmuth et al. [7], using a similar MLS approach, report 1.93 cm RMSE. All of these papers relied on LiDAR-based reconstruction and applied some form of circle-fitting in their tree modeling approaches. Additionally, the trees examined in these studies were part of human-made forest stands or naturally occurring, spatially uniform forests characterized by consistent tree structure.

In contrast to these studies, we investigated the potential of NeRF forest reconstruction combined with hull-based trunk modeling for DBH estimation in challenging terrains, featuring irregular tree structures and densely clustered tree groves. A more rigorous comparison requires evaluating methods on the same datasets. However, the datasets used in the other studies have not been made publicly available. To address this limitation and facilitate future research, we openly source our datasets for benchmarking and comparison.

The forest environment consists of harsh lighting conditions that add challenges to the use of photometric methods such as NeRFs. The dark understory created by dense forest canopy requires appropriate exposure control; long exposure times can lead to blurry images that are unusable for reconstruction purposes. A potential solution is offered by RawNeRF [27], which enables reconstruction in near-darkness environments.

The NeRF method's impressive speed-up of fieldwork time comes with the challenge of reconstructing and modeling smaller stems and complex foliage. Without optimal camera coverage, this geometry is poorly represented by the NeRF. Additionally, the filtering methods used by TreeTool need to be developed to support smaller stems and foliage. One potential avenue for improved clustering and filtering performance is to use the color of points provided by NeRF representation to aid in complex branch and foliage segmentation.

We show that convex-hull modeling is an improvement over cylinder approaches for measuring tree diameters when well-represented tree clouds are available. In practice, the presence of ground-level vegetation introduces significant occlusions in the recovered geometry. A potential solution for partial clouds could be to interpolate the cross-section from the set of stacked convex-hulls at known spacing along its height. This same slice-based modeling could be used to extract additional science measurements beyond DBH.

Parameter selection is still a semi-manual process for TreeTool. Not all parameters can be derived from density (e.g., terrain morphology dictates ground and trunk segmentation parameters). Robust, automatic parameterization can enable real-time DBH estimation in various rugged settings; further work is needed to understand the relationship between the parameter selection process and the tree-species composition of the environment.

Large-scale adoption of the methods presented in this work will require further development to support exploration volumes on the order of 100–1000 km². The NeRF scene reconstruction pipeline and corresponding computational resources described in Section 3.2 are designed to train on relatively small datasets (100–1000 images). Block-NeRF [28] provides a motivating solution to this issue by decomposing the scene into individually trained NeRFs with the ability to seamlessly align NeRFs together to reconstruct city-scale environments. The authors demonstrate the largest neural scene representation made to date, comprised of over 2.8 million training images reconstructing the neighborhoods of San Francisco, CA.

6. Conclusions

In summary, we present a field study exploring the benefits of NeRFs for DBH estimation in mixed-evergreen redwood forest. We consider an MLS comparison using LiDAR-inertial SLAM hosted on a quadruped robot. In addition, we propose a convex-hull DBH modeling technique that considerably outperformed common cylinder-fitting approaches by 3–4×. In a small-scale experiment, NeRF reconstructions made using mobile phone data outperformed SLAM in terms of DBH estimation accuracy (2.81% RMSE), at a 20× cost reduction and 5× less field-work time. In terms of relative DBH estimation accuracy, the proposed integration of NeRF with our enhanced convex-hull trunk modeling approach surpasses the performance reported in several published forest mapping studies [6,7,13,26].

While additional development is needed for autonomous ecological assessment in wilderness settings, this paper motivates the ability for rapid forest data collection using commodity mobile phone hardware. This drastic increase in accessibility has the potential of furthering community engagement, and increasing the volume of globally mapped forest terrain.

Author Contributions:

Conceptualization, A.K.; methodology, A.K.; software, A.K.; validation, A.K. and C.Y.; formal analysis, A.K.; investigation, A.K.; resources, C.J., S.M., and G.G.; data curation, A.K., C.Y., and G.G.; writing—original draft preparation, A.K.; writing—review and editing, A.K., C.J., S.M., and G.G.; visualization, A.K.; supervision, C.J.; project administration, C.J., S.M., and G.G.; funding acquisition, C.J., S.M., and G.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the ONR (grant number N00014-22-1-2290), the UCSC Center for Information Technology Research in the Interest of Society (CITRIS), and by the intramural research program of the U.S. Department of Agriculture, National Institute of Food and Agriculture. The Findings and Conclusions in This Preliminary Publication Have Not Been Formally Disseminated by the U. S. Department of Agriculture and Should Not Be Construed to Represent Any Agency Determination or Policy.

Data Availability Statement: The datasets and code contributions are freely available at: <https://github.com/harelab-ucsc/RedwoodNeRF> (accessed on 7 January 2025).

Acknowledgments: We would like to thank Morgan Masters for assisting in field data collection and providing feedback during the writing of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. EPA. *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2018*. EPA: Washington, DC, USA, 2020; pp. 430–R–20–002.
2. Davies, S.J.; Abiem, I.; Salim, K.A.; Aguilar, S.; Allen, D.; Alonso, A.; Anderson-Teixeira, K.; Andrade, A.; Arellano, G.; Ashton, P.S.; et al. ForestGEO: Understanding forest diversity and dynamics through a global observatory network. *Biol. Conserv.* **2021**, *253*, 108907.
3. Service, U.F. New aerial survey identifies more than 100 million dead trees in California. 2016. Available online: <https://www.usda.gov/media/press-releases/2016/11/18/new-aerial-survey-identifies-more-100million-dead-trees-california> (accessed on 10 June 2024).
4. Domke, G.M.; Oswalt, S.N.; Walters, B.F.; Morin, R.S. Tree planting has the potential to increase carbon sequestration capacity of forests in the United States. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 24649–24651.
5. Liang, X.; Kankare, V.; Hyypä, J.; Wang, Y.; Kukko, A.; Haggrén, H.; Yu, X.; Kaartinen, H.; Jaakkola, A.; Guan, F.; et al. Terrestrial laser scanning in forest inventories. *ISPRS J. Photogramm. Remote Sens.* **2016**, *115*, 63–77.
6. Pierzchała, M.; Giguère, P.; Astrup, R. Mapping forests using an unmanned ground vehicle with 3D LiDAR and graph-SLAM. *Comput. Electron. Agric.* **2018**, *145*, 217–225.
7. Freißmuth, L.; Mattamala, M.; Chebrolu, N.; Schaefer, S.; Leutenegger, S.; Fallon, M. Online Tree Reconstruction and Forest Inventory on a Mobile Robotic System. *arXiv* **2024**, arXiv:2403.17622.
8. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **2021**, *65*, 99–106.
9. Montoya, O.; Icasio-Hernández, O.; Salas, J. TreeTool: A tool for detecting trees and estimating their DBH using forest point clouds. *SoftwareX* **2021**, *16*, 100889.
10. Grisetti, G.; Kümmerle, R.; Stachniss, C.; Burgard, W. A tutorial on graph-based SLAM. *IEEE Intell. Transp. Syst. Mag.* **2010**, *2*, 31–43.
11. Thrun, S.; Montemerlo, M. The graph SLAM algorithm with applications to large-scale mapping of urban structures. *Int. J. Robot. Res.* **2006**, *25*, 403–429.
12. Schonberger, J.L.; Frahm, J.M. Structure-from-motion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 4104–4113.
13. Huang, H.; Tian, G.; Chen, C. Evaluating the point cloud of individual trees generated from images based on Neural Radiance fields (NeRF) method. *Remote Sens.* **2024**, *16*, 967.
14. Shan, T.; Englot, B.; Meyers, D.; Wang, W.; Ratti, C.; Rus, D. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October 2020–24 January 2021; IEEE: Piscataway, NJ, USA, 2020; pp. 5135–5142.
15. Abou-Chakra, J. NeRFCapture. Retrieved from <https://github.com/jc211/NeRFCapture>, 2023.
16. Zhang, X.; Srinivasan, P.P.; Deng, B.; Debevec, P.; Freeman, W.T.; Barron, J.T. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Trans. Graph. (ToG)* **2021**, *40*, 1–18.
17. Müller, T.; Evans, A.; Schied, C.; Keller, A. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph. (TOG)* **2022**, *41*, 1–15.
18. Barron, J.T.; Mildenhall, B.; Tancik, M.; Hedman, P.; Martin-Brualla, R.; Srinivasan, P.P. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In Proceedings of the IEEE/CVF international conference on computer vision, Montreal, BC, Canada, 11–17 October 2021; pp. 5855–5864.
19. Wang, Z.; Wu, S.; Xie, W.; Chen, M.; Prisacariu, V.A. NeRF-: Neural radiance fields without known camera parameters. *arXiv* **2021**, arXiv:2102.07064.
20. Martin-Brualla, R.; Radwan, N.; Sajjadi, M.S.; Barron, J.T.; Dosovitskiy, A.; Duckworth, D. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 7210–7219.
21. Verbin, D.; Hedman, P.; Mildenhall, B.; Zickler, T.; Barron, J.T.; Srinivasan, P.P. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 5481–5490.
22. Tancik, M.; Weber, E.; Ng, E.; Li, R.; Yi, B.; Wang, T.; Kristoffersen, A.; Austin, J.; Salahi, K.; Ahuja, A.; et al. Nerfstudio: A modular framework for neural radiance field development. In Proceedings of the ACM SIGGRAPH 2023 Conference Proceedings, Los Angeles, CA, USA, 6–10 August 2023; pp. 1–12.
23. Pingel, T.J.; Clarke, K.C.; McBride, W.A. An improved simple morphological filter for the terrain classification of airborne LIDAR data. *ISPRS J. Photogramm. Remote Sens.* **2013**, *77*, 21–30.
24. Schubert, E.; Sander, J.; Ester, M.; Kriegel, H.P.; Xu, X. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Trans. Database Syst. (TODS)* **2017**, *42*, 1–21.

25. Gilbert, G.S.; Carvill, S.G.; Krohn, A.R.; Jones, A.S. Three Censuses of a Mapped Plot in Coastal California Mixed-Evergreen and Redwood Forest. *Forests* **2024**, *15*, 164.
26. Liang, X.; Kankare, V.; Yu, X.; Hyypä, J.; Holopainen, M. Automated stem curve measurement using terrestrial laser scanning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 1739–1748.
27. Mildenhall, B.; Hedman, P.; Martin-Brualla, R.; Srinivasan, P.P.; Barron, J.T. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 16190–16199.
28. Tancik, M.; Casser, V.; Yan, X.; Pradhan, S.; Mildenhall, B.; Srinivasan, P.P.; Barron, J.T.; Kretschmar, H. Block-nerf: Scalable large scene neural view synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 8248–8258.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.