*remote sensing*

MDPI

*Article*

# Evaluating the Point Cloud of Individual Trees Generated from Images Based on Neural Radiance Fields (NeRF) Method

**Hongyu Huang [1,2,3,*], Guoji Tian [1,2,3] and Chongcheng Chen [1,2,3]**

[1] National Engineering Research Center of Geospatial Information Technology, Fuzhou University, Fuzhou 350108, China; 225527044@fzu.edu.cn (G.T.); chencc@fzu.edu.cn (C.C.)
[2] Key Laboratory of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University, Fuzhou 350108, China
[3] The Academy of Digital China (Fujian), Fuzhou 350108, China
* Correspondence: hhy1@fzu.edu.cn; Tel.: +86-13328269460

**Abstract:** Three-dimensional (3D) reconstruction of trees has always been a key task in precision forestry management and research. Due to the complex branch morphological structure of trees themselves and the occlusions from tree stems, branches and foliage, it is difficult to recreate a complete three-dimensional tree model from a two-dimensional image by conventional photogrammetric methods. In this study, based on tree images collected by various cameras in different ways, the Neural Radiance Fields (NeRF) method was used for individual tree dense reconstruction and the exported point cloud models are compared with point clouds derived from photogrammetric reconstruction and laser scanning methods. The results show that the NeRF method performs well in individual tree 3D reconstruction, as it has a higher successful reconstruction rate, better reconstruction in the canopy area and requires less images as input. Compared with the photogrammetric dense reconstruction method, NeRF has significant advantages in reconstruction efficiency and is adaptable to complex scenes, but the generated point cloud tend to be noisy and of low resolution. The accuracy of tree structural parameters (tree height and diameter at breast height) extracted from the photogrammetric point cloud is still higher than those derived from the NeRF point cloud. The results of this study illustrate the great potential of the NeRF method for individual tree reconstruction, and it provides new ideas and research directions for 3D reconstruction and visualization of complex forest scenes.

**Keywords:** 3D reconstruction; neural radiance field (NeRF); 3D tree modeling; photogrammetry; deep learning; individual tree; terrestrial laser scanning; lidar



**Citation:** 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. https://doi.org/10.3390/rs16060967

Academic Editor: Devrim Akca

Received: 7 February 2024
Revised: 7 March 2024
Accepted: 7 March 2024
Published: 10 March 2024



## 1. Introduction

Trees are an essential part of the Earth's ecosystem, providing numerous critical ecological services and influencing many environmental aspects such as soil conservation, climate regulation and wildlife habitats [1]. A comprehensive understanding of the distribution and number of trees, as well as information on tree morphology and structure, is important for forestry and natural resource management, as well as environmental monitoring, characterization and protection. The rapid development of 3D digitization, reconstruction technology and artificial intelligence (AI) provides a new direction for tree monitoring and protection. Three-dimensional technology is applied to forestry by collecting data and then performing automatic modeling, and key information such as tree height, diameter at breast height (DBH), crown diameter and volume of trees can be obtained for forest inventory and management purposes [2].

To acquire 3D data of trees, a variety of means can be used, which can usually be categorized into terrestrial platforms and aerial platforms. Ground-based platforms include terrestrial laser scanning (TLS), mobile laser scanning (MLS) and ground-based photogrammetry [3]. TLS has the highest geometric data quality of all sensors and platforms, while

MLS collects data efficiently but with slightly less accuracy than TLS. Ground-based photogrammetry, on the other hand, processes photos collected from a variety of cameras, which is commonly used and easy to operate. Airborne platforms mainly include UAVs, helicopters, etc., which are capable of acquiring high spatial resolution data quickly, with accuracy even comparable to that of ground-based acquisition systems.

Terrestrial laser scanners are often used for forest surveys and 3D tree reconstruction, as they are able to acquire accurate point cloud models of trees in a non-contact, active manner. However, TLS is relatively expensive and difficult to carry around, which prevents it from being used easily in dense forests. In the last decade, the photogrammetric method that is simple and affordable to use for acquiring 3D data has become widely used for forest inventory and vegetation modeling works [4]. Huang et al. used a consumer-grade handheld camera to acquire images of a desert plant and were able to reconstruct a point-cloud model with accuracy comparable to that of from TLS [5]; Kükenbrink et al. [6] evaluated the performance of an action camera (GoPro) combined with the Structure from Motion (SfM) [7] technique to acquire 3D point clouds of trees, and their point cloud model was of similar quality to that of Lidar devices (TLS, MLS and UAV-laser scanning) in terms of the DBH extraction results, but still fell short of the performance in other aspects. For this reason, many scholars have opened up new ideas to acquire forest point cloud data using a portable camera and a backpack laser scanner, combining these two types of data to extract tree height and DBH [8]; Balestra et al. fused the data from a RGB camera, UAV camera and mobile laser scanner to reconstruct three giant chestnut tree models, providing researchers with accurate information about the shape and overall condition of the trees [9]. Although these photogrammetry-based 3D tree reconstruction methods are simple, automated and effective, the reconstruction accuracy is greatly affected by the fact that the trees themselves have complex self-similar branch morphological structures and the inner branches are shaded by leaves. The current photogrammetry method can deal with trees with prominent stem and branch features that are not covered or concealed by the foliage, but for trees with dense leaves in their canopies, the reconstruction result is often not satisfactory: usually only the lower trunk or stem can be recovered, while the upper canopy is partly or totally missing in the 3D model. New methods that can reconstruct complete 3D tree canopies accurately from images are highly anticipated and appreciated.

After the concept of deep learning [10] was first introduced in 2006, it has been developed rapidly in various fields, including computer graphics and 3D reconstruction. Neural Radiance Fields (NeRF) [11] is one example of a recent major development. NeRF create photorealistic 3D scenes from series photos. Given the input of a set of calibrated images, the goal is to output a volumetric 3D scene that renders novel views. NeRF's main task is to synthesize new views based on the known view images, but can also be used to generate mesh using marching cubes [11]; further, in nerfstudio (https://docs.nerf.studio/index.html, accessed on 22 November 2023), the reconstructed mesh and point cloud can be exported. Since the original NeRF was proposed, the field has grown explosively with hundreds of papers extending or building on it each year, and this method has found new applications in various areas, including autonomous driving [12], medicine [13], digital human body [14], 3D cities [15] and cultural heritage reconstruction [16,17], to name a few.

A few studies have been conducted to compare the relatively mature photogrammetry reconstruction method to NeRF to understand their advantages and disadvantages. The authors of [18] provided a comprehensive overview and analysis of the strengths and weaknesses of NeRF and photogrammetry and evaluated the quality of the generated 3D data on objects of different sizes and surface properties. The results show that NeRF outperforms photogrammetry on texture-less, metallic, highly reflective and transparent objects, but photogrammetry still performs better with textured objects [18]. The objects studied included statue, truck, stair and bottles, which are man-made, static objects; however, there has been no published research on NeRF for 3D reconstruction of living, natural objects such as trees so far. Therefore, in this paper, we applied NeRF technology to achieve 3D reconstruction of trees and used the derived point cloud for tree structural parameters

extraction and 3D modeling. Our goal is to answer these questions: how good is the derived point cloud (both visually and quantitatively) compared with traditional photogrammetric dense reconstruction methods and what are the strengths and weaknesses of this method for 3D vegetative modeling? We wish to use this study to collect information and gain new ideas and directions for the 3D reconstruction of trees.

## 2. Materials and Methods

### *2.1. Study Area*

We chose two trees with unique features located within the Qishan Campus of Fuzhou University as representative examples for this study. Tree_1, a recently transplanted Autumn Maple tree, is located in a fairly open area; it was easy to take photos around the tree as the view was not being blocked or interfered by other vegetative materials. The tree has distinctive trunk characteristics, sparse foliage and a relatively simple crown and branch structure. Tree_2 is an imposing Acacia tree that is situated in a crowded, densely vegetated area, with several big trees close by. This presented a challenge to fully or adequately take sample images of the tree. Tree_2 is tall, with a wide-spread canopy, dense foliage and complex branch structure. Figure 1 shows the morphology of these two trees.

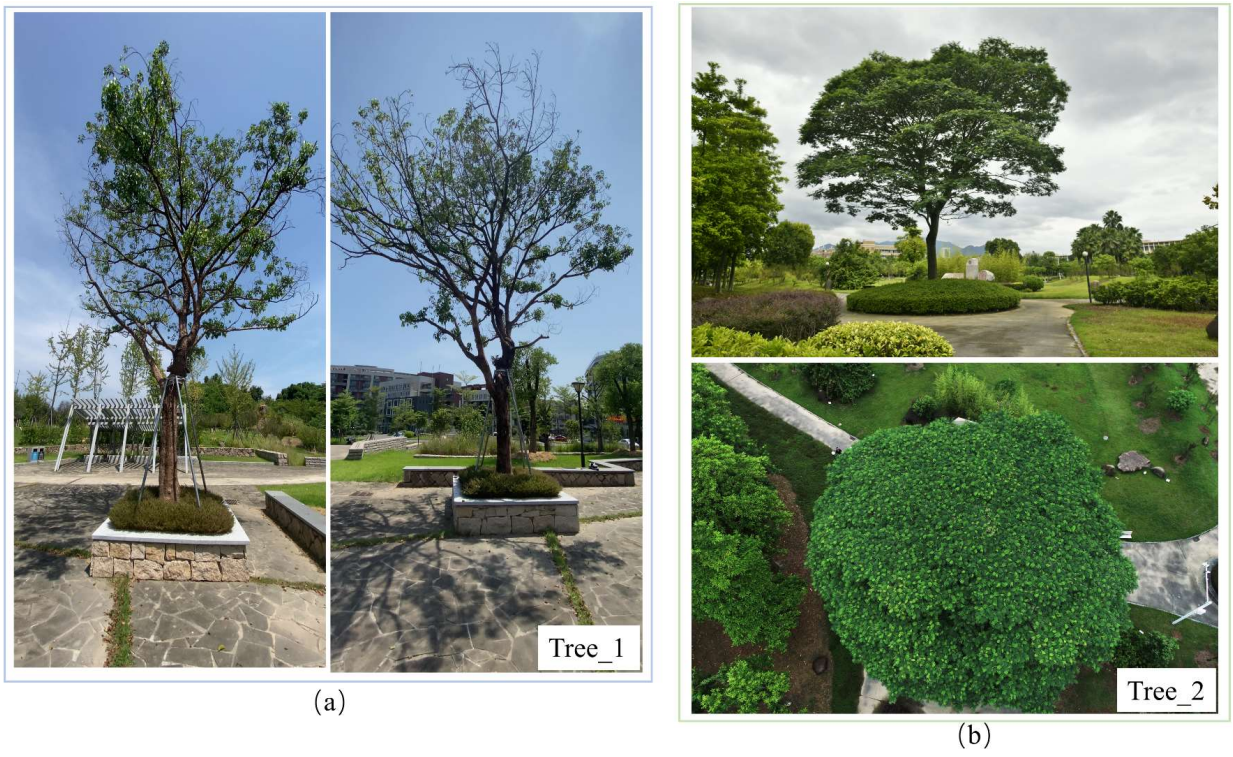


**Figure 1.** The structures and shapes of two trees used in this study. (**a**) Tree_1 with views from two perspectives on the ground and (**b**) Tree_2 with views from both the ground and in the air.

### *2.2. Research Methods*

#### 2.2.1. Traditional Photogrammetric Reconstruction

The development of photogrammetric reconstruction based on multiple overlapping images is relatively mature by now, in which Structure from Motion (SfM) [7] and multi-view stereo (MVS) [19] are the two major processing steps. SfM originates from computer vision; it is the process of reconstructing 3D structure of the scene from a series of images taken from different viewpoints (motions of the camera). The results of SfM are the exterior and interior parameters of the camera and sparse feature points of an object or scene. MVS, on the other hand, is designed to reconstruct a complete, dense 3D object model from a collection of images taken from known camera viewpoints.

This typical reconstruction process is implemented in many open source and commercial software including, but not limited to, COLMAP (https://colmap.github.io/, accessed on 24 November 2023), Agisoft Metashape (https://www.agisoft.com/, accessed on 22 November 2023) and Pix4Dmapper (https://www.pix4d.com/pix4dmapper, accessed on 20 November 2023), among many others. We used the open-source COLMAP (Version 3.6), a popular and established general-purpose SfM and MVS pipeline, as the photogrammetric tool for the reconstruction comparison experiments; NeRF also uses COLMAP's SfM package to generate calibrated images for neural network training.

#### 2.2.2. Neural Radiance Fields (NeRF) Reconstruction

Neural Radiance Fields (NeRF) is able to implicitly represent a static object or scene with a multilayer perceptron (MLP) neural network, which is optimized to generate a picture of the scene from any perspective [11]. NeRF's superb ability to represent implicit 3D information has led to its rapid application in areas such as new perspective synthesis and 3D reconstruction. The general workflow of NeRF is shown in Figure 2. The principle is to represent a continuous scene implicitly by a function whose input is a 5D vector (3D position coordinates X = ($x$, $y$, $z$) and 2D view direction d = ($\theta$, $\phi$)), and the output is the color information $c$ = (r, g, b) and the bulk density σ about the point at X. In NeRF, the function is approximated with a MLP continuously optimized for the implementation. This function can be denoted as

$$F_\theta: (\mathrm{X}, \mathrm{d}) \rightarrow (c, \sigma) \quad (1)$$

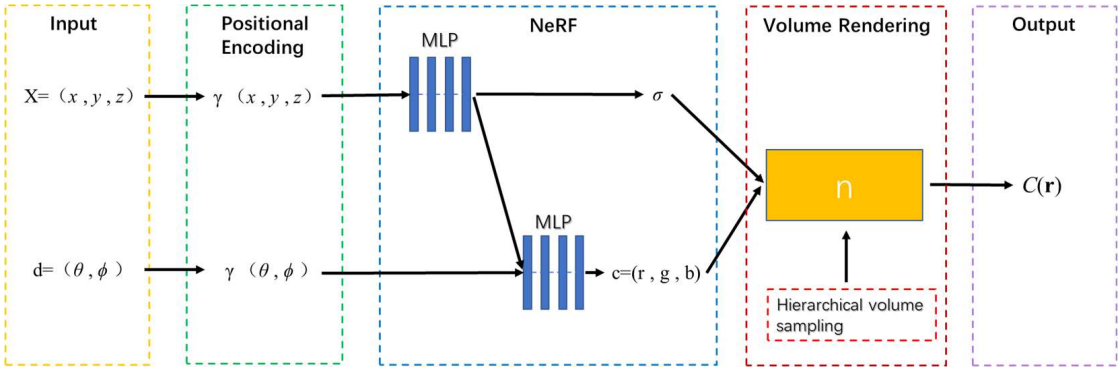


**Figure 2.** Overview of neural radiance field scene representation. Images are synthesized by sampling 5D coordinates (location and viewing direction) along camera rays. After mapping this position information to high-dimensional space through position encoding, they are then fed into MLP to generate color and volume density. Volume rendering techniques are used to composite these values into an image.

In order to speed up the convergence of the MLP network, NeRF also proposes Positional Encoding, which uses a high-frequency mapping function γ to map the coordinates and viewing direction to a high-dimensional space to obtain the coordinate encoding γ (x) and the direction encoding γ (d).

The volume rendering method is then used to render the density and color information obtained from the MLP. The volume density can be understood as the different probabilities that a ray terminates in an infinitesimal particle. Therefore, the volume density and color can be integrated along the ray, with N uniform sampling points, the cumulative transmittance from near to far is used as the integral weight, and then the volume rendering result is obtained. That is, color $C$ can be expressed as

$$C = \sum_{i=1}^{N} T_i (1 - exp(-\sigma_i \delta_i)) c_i \quad (2)$$

$$T_i = exp\left(-\sum_{j=1}^{i-1} \sigma_i \delta_i\right) \quad (3)$$

where $\delta_i$ denotes the distance between the consecutive samples ($i$ and $i + 1$), while $\sigma_i$ and $c_i$ represent the estimated density and color values along the sample point ($i$) and $T_i$ denotes cumulative transmittance.

In addition, in order to reduce the impact of excessive calculations caused by too many sampling points on the ray, NeRF adopts a hierarchical sampling method from "coarse" to "fine", which can allocate computing resources effectively. And the parameters of the MLP network are optimized by the Loss function between the volume-rendered synthetic image and the real image.

Although the initial NeRF method is concise and effective, it also suffers from problems such as long training time and aliasing artifacts. In order to solve these problems and improve the performance of the NeRF method, many new methods have been developed: the Instant-ngp [20] and Mip-NeRF [21] are two such notable examples. The most significant recent advancement is nerfstudio [22], which is a modular PyTorch framework for NeRF development that integrates various NeRF methods. Nerfstudio allows users to train their own NeRFs with some basic coding ability and knowledge and suitable hardware. It provides a convenient web interface (viewer) to display the training and rendering process in real-time and can export the rendering results to video, point cloud and mesh data. Nerfstudio's default and recommended method is Nerfacto [23], which draws on the strengths of several methods to improve its performance. It uses a piecewise sampler to produce the initial set of samples of the scene, which makes it possible to sample even distant objects. These samples are then input to the Proposal Sampler proposed in MipNeRF-360 [24], which consolidates the sample locations to the regions of the scene that contribute most to the final render (typically the first surface intersection). In addition, Nerfacto combines a hash encoding with a small fused MLP (from Instant-ngp) to realize the density function of the scene, which ensures accuracy and high computational efficiency.

In summary, we used the nerfstudio framework for the NeRF reconstruction experiment.

### *2.3. Data Acquisition and Processing*

#### 2.3.1. Data Acquisition

Several tools were used for data collection. A smartphone camera was used to take photos and record video around the trees on the ground. In addition, for Tree_2, a Nikon digital camera was also used to collect photos on the ground, and a DJI Phantom 4 UAV was used to acquire image data both on the ground and in the air. These different types of consumer-grade cameras are accessible and widely used; the image resolution and numbers of acquired images are shown in Table 1. Considering the need for quality ground truth data, a RIEGL VZ-400 terrestrial laser scanner (main specifications: ranging accuracy 3 mm, precision 5 mm, laser beam divergence 0.35 mrad) was used to perform multi-station scanning of the two target trees to obtain point cloud data for reference. Tree_1 and Tree_2 was scanned from 3 and 4 stations, respectively, with a scanning angular resolution of 0.04 degrees. The data were fine-registered in RiSCanPro (http://www.riegl.com/products/software-packages/riscan-pro/, accessed on 2 July 2023) with an accuracy of about 5 mm.

**Table 1.** Image data information sheet.

| Image Dataset | Number of Images | Image Resolution | Total Pixel (Millions) |
|---|---|---|---|
| Tree_1_phone | 118 | 2160 × 3840 | 979 |
| Tree_2_phone | 237 | 1080 × 1920 | 491 |
| Tree_2_nikon | 107/66 | 8598 × 5597 | 5149 |
| Tree_2_uav | 374 | 5472 × 3648 | 7466 |

Note: The first part of the dataset type name represents the target tree, and the last part describes the image collection sensor, which can be smartphone camera (phone), digital camera Nikon (nikon) or drone (uav). For example, Tree_2_uav refers to the data collected using a drone on Tree_2. For Tree_1_phone and Tree_2_phone, 118 and 237 frames were extracted from the recorded videos. For Tree_2_nikon, there were 107 images in total, but only 66 of these images could be calibrated after the SfM procedure.

2.3.2. Data Processing

Photogrammetry and NeRF share the same first step of processing, which is SfM, whose results include recovered image position and orientation, as well as sparse feature points of the scene. Multi-View Stereo (MVS) in COLMAP uses the output of SfM to compute depth and/or normal information for each pixel in an image. Merging the depth and normal maps from multiple 3D images then produces a dense point cloud of the scene. For NeRF the images and their positions and orientations were used for training and validating the MLP model. Object and scene points can be generated and exported from nerfstudio; we then compared three sets of tree point clouds for their visual appearances and information contents after bringing both point clouds derived from the photogrammetry and NeRF methods to the common coordinate system of the TLS point cloud and changed their scales. Figure 3 shows the workflow of this study. All the image processing (photogrammetry and NeRF) was conducted in the same configuration setting of a cloud server platform; the computing equipment is equipped with Windows10 operating system, 12-core CPU, 28 GB of RAM and NVIDIA GeForce RTX 3090 (24 GB VRAM) GPU.

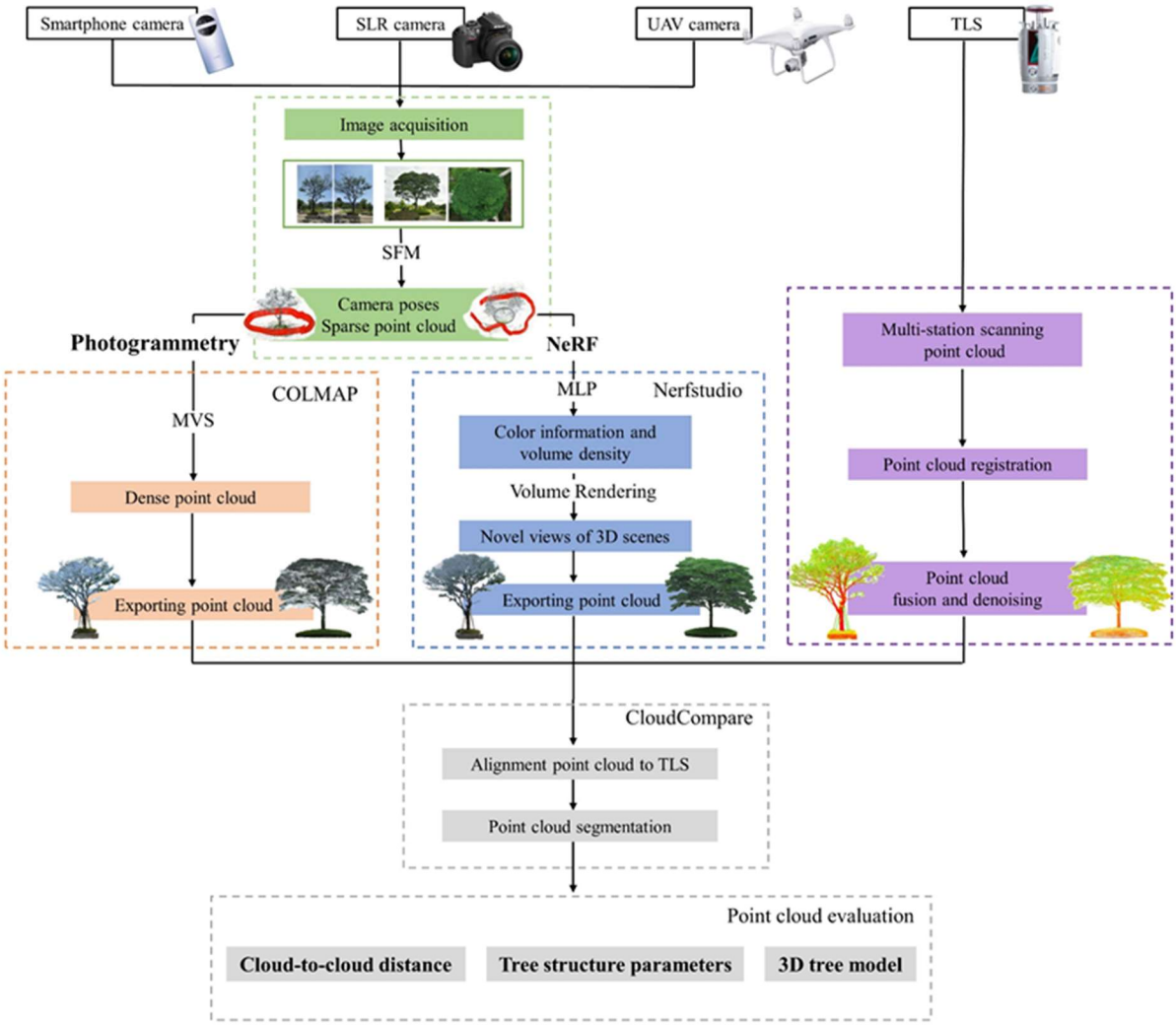


**Figure 3.** Complete workflow of the experiment. Images were acquired by different cameras from various perspectives; they were then processed using Structure from Motion (SFM) to derive camera poses and sparse point cloud. For dense point cloud generations, two methods were tested: Multi-View Stereo (MVS) in COLMAP and NeRF in Nerfstudio. These dense point clouds were finally compared to the reference point cloud obtained from terrestrial laser scanning (TLS).

## 3. Results

### 3.1. Reconstruction Efficiency Comparison

Since SfM is a step utilized by both the NeRF and photogrammetry methods, in a sense NeRF is similar to the MVS step in photogrammetry for generating dense points of the scene. So we compare the time used for COLMAP to carry out the densification process and the time the NeRF method used for training a 10,000-epoch neural network. We observed that after 7000 epochs of training, the Loss function usually converged, so 10,000 epochs seems to be a conservative number. As shown in Table 2, COLMAP spent much more time than NeRF did to produce dense points of the scene. The time consumed by NeRF reconstruction always stayed within the range of 10–15 min for different image sequences, while the time taken by COLMAP was 4 to 9 times longer and it increased with the number of images. For the Tree_2_uav dataset, the COLMAP program failed to generate dense points after running for more than 127 min due to internal error.

**Table 2.** Computation time (minutes) of dense reconstruction for different trees and methods.

| | Tree_1_phone | Tree_2_phone | Tree_2_nikon | Tree_2_uav |
|---|---|---|---|---|
| COLMAP | 98.003 | 102.816 | 50.673 | failed |
| NeRF | 11.5 | 12.0 | 12.5 | 14.0 |

### 3.2. Point Cloud Direct Comparison

We imported the point clouds obtained from COLMAP and NeRF reconstruction into CloudCompare (https://www.cloudcompare.org/, accessed on 2 July 2023) and finely aligned them with the TLS Lidar point cloud. We then extracted the targeted trees from the scene manually before conducting further comparative analyses. For each reconstructed tree point cloud model, we calculated its cloud-to-cloud distance to the corresponding TLS point cloud using CloudCompare.

As shown in Table 3, in most cases more than 1 million points could be reconstructed from images for each tree's 3D point cloud model, and COLMAP produced more points than the corresponding NeRF. It also shows that COLMAP was not able to deal with the images taken by the Nikon for Tree_2 very well, indicated by the low number of points for Tree_2_nikon_COLMAP. And as mentioned in the previous section, COLMAP failed to densely reconstruct Tree_2_uav.

**Table 3.** Number of points of the tree point cloud models.

| Tree ID | Model ID | Number of Point |
|---|---|---|
| Tree_1 | Tree_1_Lidar | 990,265 |
| | Tree_1_COLMAP | 1,506,021 |
| | Tree_1_NeRF | 1,275,360 |
| Tree_2 | Tree_2_Lidar | 2,986,309 |
| | Tree_2_phone_COLMAP | 1,746,868 |
| | Tree_2_phone_NeRF | 1,075,874 |
| | Tree_2_nikon_COLMAP | 580,714 |
| | Tree_2_nikon_NeRF | 1,986,197 |
| | Tree_2_uav_NeRF | 1,765,165 |

Note: The first part of the Model ID name represents the name of the target tree, the middle part describes the image collection sensor, which can be smartphone camera (phone), digital camera Nikon (nikon) or drone (uav), and the last part denotes the image reconstruction method, which can be either COLMAP or NeRF. For Tree_1, the only image sensor is the phone, so it is neglected from the name.

Figure 4 presents the three versions of point cloud models of Tree_1 and illustrates their differences. COLMAP and NeRF both reconstructed the target tree's trunk, canopy and the surrounding ground scene with color completely, but both have white noise in their canopies. COLMAP's noise is distributed on the crown and trunk surfaces, while NeRF's

is concentrated on the crown surface. To quantify the spatial distribution of the points, we calculated the cloud-to-cloud (C2C) distance after SOR denoising in Cloud Compare. The C2C distance refers to the measurement of the nearest neighboring distance between corresponding points in two point clouds.

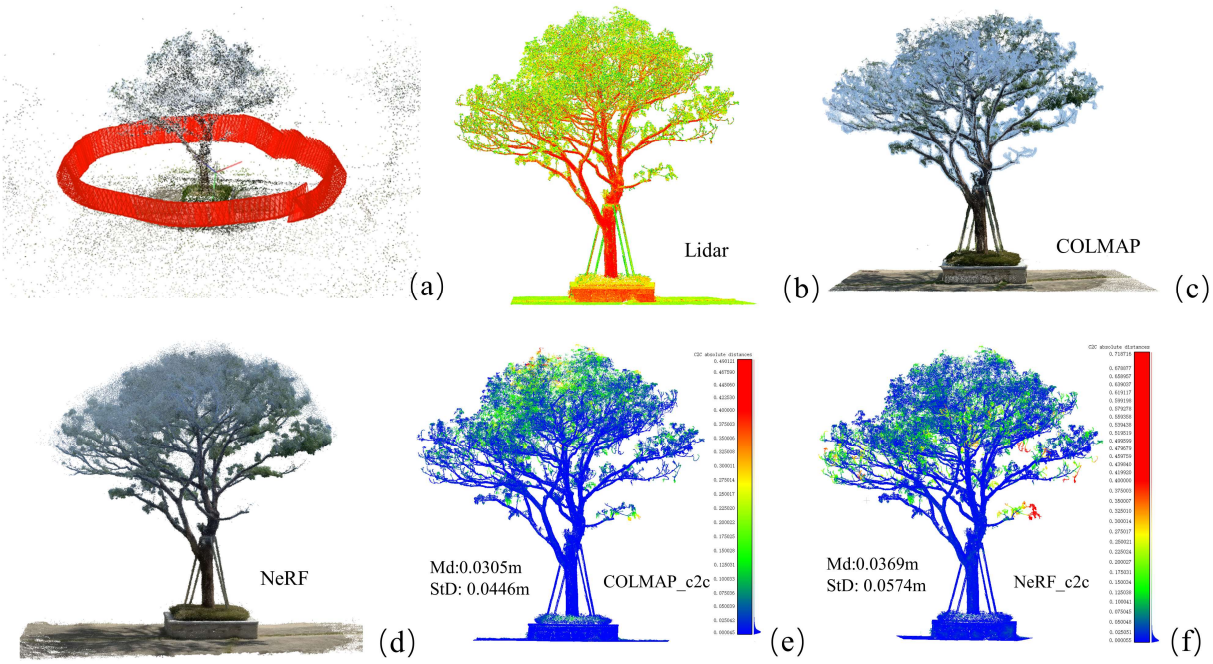


**Figure 4.** Tree_1 and its reconstruction result comparisons: (**a**) camera poses shown in red and sparse points of the scene; (**b**) TLS Lidar point cloud, with intensity values colored in red and green representing trunks (branches) and leaves; (**c**) COLMAP model with color in RGB; (**d**) NeRF model, also color in RGB; (**e**) the cloud-to-cloud (c2c) distance between TLS Lidar and COLMAP model; (**f**) the c2c distance between TLS and NeRF model. Md is mean distance, StD is standard deviation.

We report mean distance (*Md*) and standard deviation (*StD*):

$$Md = \frac{\sum_{i=1}^{n} x_i}{n} \tag{4}$$

$$StD = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}} \tag{5}$$

where $n$ denotes the number of points involved in the distance comparison, $X_i$ denotes the distance from Point $i$ to its nearest neighbor and $\overline{X}$ denotes the mean distance.

As shown in Figure 4e,f, in the color scale bar, we set all the distances greater than 0.4 m to be red. It can be seen that relative to the Lidar data, COLMAP has missing or distant tree parts at the top of the canopy, while the NeRF has missing or distant tree elements in the middle of the canopy and at some branches. In addition, the mean cloud distance of the Lidar model with respect to the COLMAP and NeRF models are 0.0305 m and 0.0369 m, with standard deviations of 0.0446 m and 0.0574 m, respectively.

As shown in Figure 5 and Table 3, among the six versions of the Tree_2 point cloud model, the TLS Lidar model has the largest number of points and the best quality. It is followed by the NeRF_uav model, which better reproduces the true color of the tree and has very little canopy noise. The COLMAP_nikon model has the worst quality, which only reconstructs the trunk portion of the tree, with the outer canopy shell partially recreated and the more complex canopy interior portion missing. For COLMAP_phone and NeRF_phone, the models both successfully reconstructed these from image frames taken from a smart phone video; the reconstruction quality was better even though the number of points was

700,000 less for the NeRF method, which can reconstruct more dendritic information within the canopy with less noise. Similarly, we took the point cloud models of COLMAP and NeRF as a reference and calculated the cloud-to-cloud distance between them and the TLS model. Then, in the scale bar, we set the areas with distances greater than 1 m to be red and assumed that these parts were most likely to be missing. It can be seen that the NeRF_uav and NeRF_nikon models are the best in quality when compared to the LiDAR data, with only some missing branches in the lower part of the canopy, while the worst is the COLMAP_nikon model, with almost all missing branches and leaves inside the canopy. In addition, the mean cloud distances of the LiDAR model with NeRF_uav, NeRF_nikon and COLMAP_nikon are 0.0644 m, 0.0747 m and 0.3035 m, with standard deviations of 0.0915 m, 0.1127 m and 0.3160 m, respectively. The NeRF_phone model has more branches inside the canopy compared to the COLMAP_phone model, although some branches are missing in the lower part of the canopy, and the COLMAP_phone model has more missing elements inside the canopy. The mean cloud distances for NeRF_phone and COLMAP_phone are 0.1038 m and 0.1453 m, with standard deviations of 0.1539 m and 0.1922 m, respectively.

### *3.3. Extraction of Structural Parameters from Tree Point Cloud*

We extracted the tree structural parameters from the point cloud derived from TLS and reconstructed from images via photogrammetry and NeRFs using commercial and open source software Lidar360 (Version 6.0, https://www.lidar360.com/, accessed on 2 September 2023) and 3DForest (Version 0.5, https://www.3dforest.eu/, accessed on 2 September 2023), respectively. The parameters derived from the laser scanning point cloud model were used as a reference to compare with those derived from different reconstructed models.

The structural parameters of tree height (TH), diameter at breast height (DBH), crown diameter (CD), crown area (CA) and crown volume (CV) of a single tree were obtained after a series of processing steps such as denoising and single-tree segmentation in Lidar360 software, as summarized in Table 4. Due to limited or incomplete reconstructed canopy, Tree_2_nikon_COLMAP was not involved in the following processing.

**Table 4.** Results of Lidar360 extracted structural parameters for the studied trees.

| Models | TH (m) | DBH (m) | CD (m) | CA ($m^2$) | CV ($m^3$) |
|---|---|---|---|---|---|
| Tree1_Lidar | 8.2 | 0.345 | 7.1 | 39.5 | 137.0 |
| Tree1_NeRF | 8.4 | 0.318 | 7.1 | 39.2 | 139.7 |
| Tree1_COLMAP | 8.1 | 0.349 | 7.0 | 38.9 | 138.0 |
| Tree2_Lidar | 13.6 | 0.546 | 16.3 | 208.8 | 1549.2 |
| Tree2_uav_NeRF | 14.0 | 0.479 | 16.0 | 201.0 | 1531.9 |
| Tree2_nikon_NeRF | 14.7 | 0.461 | 16.5 | 214.6 | 1627.0 |
| Tree2_phone_NeRF | 14.3 | 0.469 | 17.0 | 225.7 | 1693.1 |
| Tree2_phone_COLMAP | 13.8 | 0.562 | 16.7 | 220.3 | 1657.4 |

From Table 4, it can be seen that the structural parameters of both the COLMAP and NeRF models for Tree_1 are not much different from those derived from the Lidar model, used as ground truth here, with errors staying in a narrow range. The COLMAP-extracted parameters are closer to the Lidar-derived values in tree height, DBH and crown volume metrics, but the NeRF-extracted values have smaller errors in crown diameter and crown area, which is also in line with the results of the visual comparisons. For Tree_2, both the COLMAP and NeRF models have higher tree height estimates than those of the Lidar model, with phone_COLMAP being the one with the least errors. In terms of DBH, again in agreement with the results of the visual comparisons, the DBH of the NeRF tree model was smaller than that of the Lidar model with a range between 6.7 cm and 8.5 cm, whereas the DBH of the COLMAP model was coarser than the Lidar model by 1.6 cm. And the errors of crown diameter were uniformly in the range of 0.2 m to 0.7 m. The uav_NeRF

is closer to Lidar in terms of crown area and volume, and the model reconstructed using smartphone images deviates most from Lidar's values.

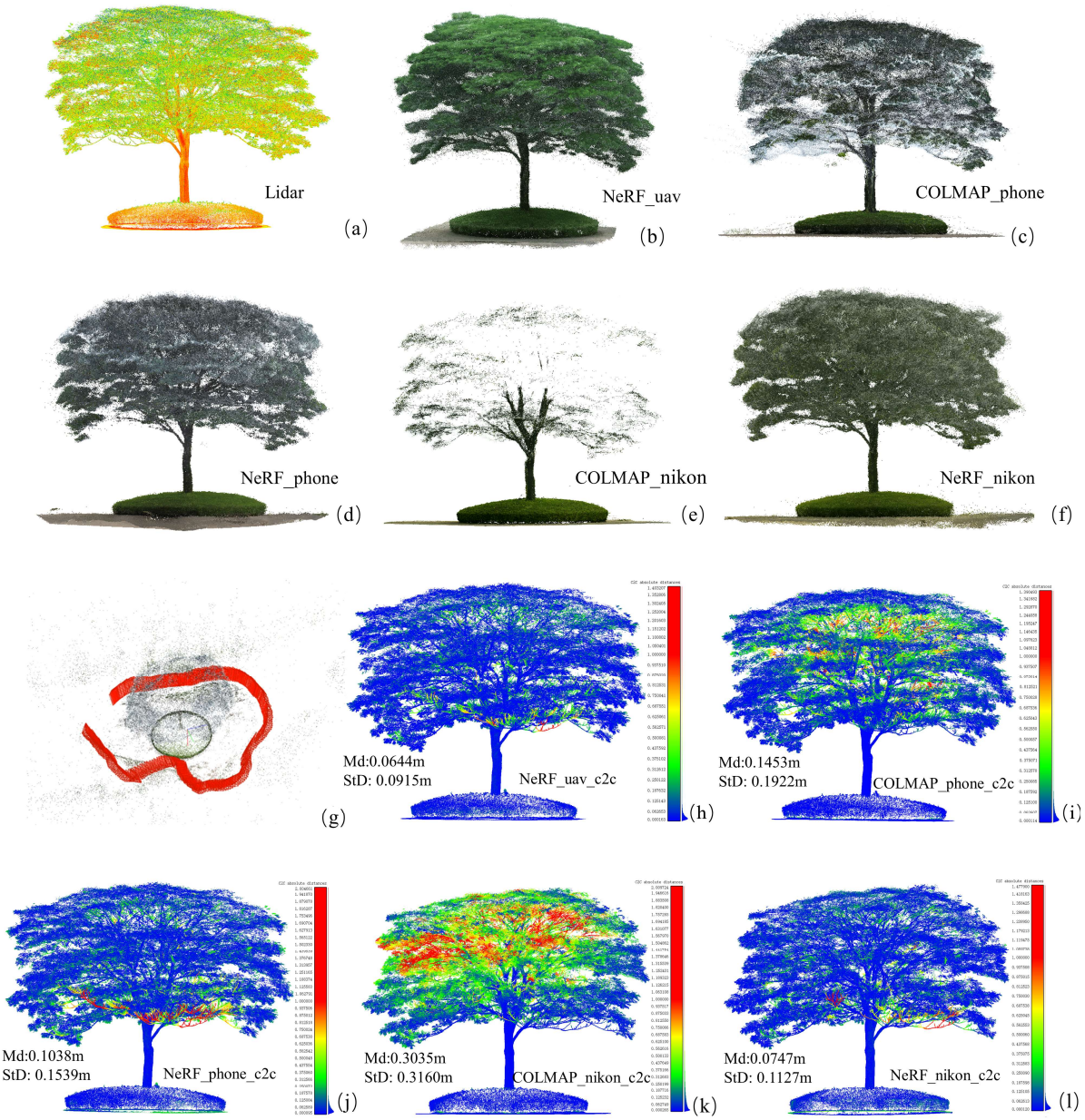


**Figure 5.** Tree_2 and its reconstruction result comparisons: (**a**) TLS Lidar point cloud, with intensity values colored in red and green representing trunks (branches) and leaves; (**b**) NeRF_uav model with color in RGB; (**c**) COLMAP_phone model with color in RGB; (**d**) NeRF_phone model with color in RGB; (**e**) COLMAP_nikon model with color in RGB; (**f**) NeRF_nikon model, also color in RGB; (**g**) camera poses of the smartphone shown in red and sparse points of the scene; (**h**) the cloud-to-cloud (c2c) distance between TLS and NeRF_uav model; (**i**) the c2c distance between TLS and COLMAP_phone model; (**j**) the c2c distance between TLS and NeRF_phone model; (**k**) the c2c distance between TLS and COLMAP_nikon model; (**l**) the c2c distance between TLS and NeRF_nikon model. Md is mean distance, StD is standard deviation.

To compare different algorithms and test the usability of the derived point cloud, we also used 3DForest to extract certain important structural parameters, namely, the tree height, DBH and tree length (TL). The results are shown in Table 5:

**Table 5.** Extraction of major structural parameters of the models by 3DForest.

| Models | TH (m) | DBH (m) | TL (m) |
|---|---|---|---|
| Tree_1_Lidar | 8.16 | 0.347 | 8.07 |
| Tree_1_NeRF | 8.32 | 0.321 | 7.95 |
| Tree_1_COLMAP | 8.02 | 0.346 | 8.04 |
| Tree_2_Lidar | 13.35 | 0.548 | 17.20 |
| Tree_2_uav_NeRF | 13.75 | 0.476 | 17.19 |
| Tree_2_nikon_NeRF | 13.55 | 0.468 | 18.0 |
| Tree_2_phone_NeRF | 13.89 | 0.502 | 18.22 |
| Tree_2_phone_COLMAP | 13.86 | 0.557 | 17.64 |

In Table 5, among the single tree model parameters extracted by 3DForest, DBH and TH remained consistent with those extracted by Lidar360, with the difference on the centimeter or even millimeter scale, which proved the reliability and accuracy of the extracted model parameters, as well as the quality of the input data and robustness of the relevant processing algorithms.

From these structural metrics extracted from both Lidar360 and 3Dforest, it appears that from the NeRF point cloud model of trees that the tree height tends to be over-estimated, while the DBH is under-estimated. This feature may be related to how the meshes and points were derived from the underlying volumetric scene representation.

### *3.4. Comparison of 3D Tree Models Generated from Point Cloud*

Based on the Lidar, COLMAP and NeRF reconstruction point cloud models, tree modelling was performed using the AdTree [25] open source project to generate 3D models of the target trees. AdTree uses a minimum spanning tree (MST) algorithm to efficiently extract the initial tree skeleton on the input tree point cloud and then obtains a tree model with reconstructed branches through iterative skeleton simplification and cylinder fitting. Based on these 3D models, further comparative analyses can be performed to explore the usability and accuracy of point clouds generated by different reconstruction methods.

Figure 6 shows point cloud models, generated 3D branch models and the superimposed point cloud on branch models for detailed comparison. At first glance, all three branch models bear a high degree of resemblance to their corresponding point cloud models; but closer examination reveals more or less disconformities for each of them. Lidar point cloud has the best quality, but there are some branches that have not been reconstructed (marked in the blue box); NeRF's branch model looks messy in the tree crown, with a lot of smaller branches; and COLMAP's crown looks better than that of NeRF's, but in COLMAP's branch model there is some distortion at the trunk, while NeRF's trunk is smoother (marked in yellow box). The merged point and branch model reveals that the trunks of all three branch models are a bit thicker than the point cloud models (marked in red boxes).

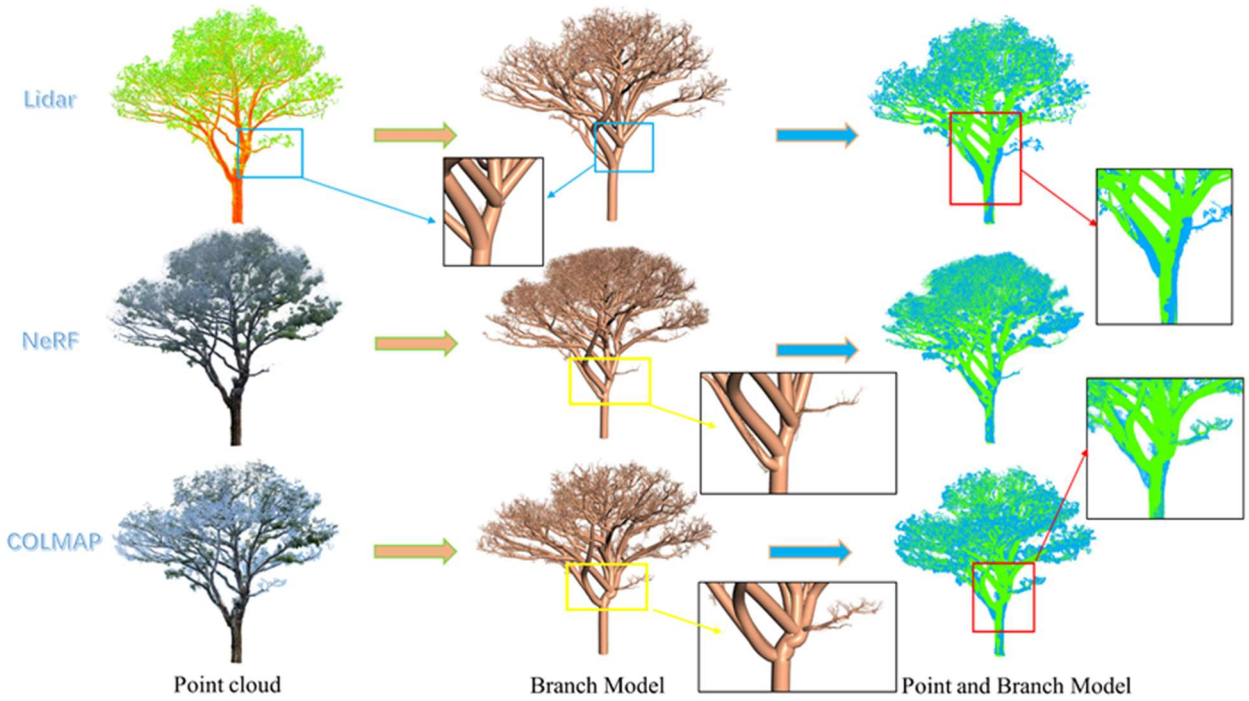


**Figure 6.** Tree_1 and its 3D branch models generated by AdTree. Left column is point cloud model, with the LiDAR model colored with intensity values and the NeRF and COLMAP models colored in RGB; middle column is branch model; right column is merged point and branch model, with the point cloud model color in blue and the branch model in green. The color boxes are used to highlight the details of the model.

The AdTree-generated 3D branch models for Tree_2 are shown in Figure 7. Overall, the tree height and tree morphology of the branch models are consistent with the point cloud models, except for the COLMAP_phone, which has substantial differences with the point cloud model in the canopy part. These upright-growing branches in the crown (marked in green box) are not found in the point cloud model. In addition, the trunk position and shape of the Lidar, NeRF_uav and NeRF_nikon models are similar, while those of NeRF_phone and COLMAP_phone are different (marked in blue boxes). NeRF_phone's trunk is misplaced from its position in the point cloud, while the latter's trunk is simply incorrect. Merged point and branch models demonstrate that the trunk and branch portion of the branch model is thicker than the point cloud model, where Lidar and NeRF_uav's models are closer to the point cloud model, while the COLMAP_phone model has the largest deviation from the point cloud model (marked in the yellow box).

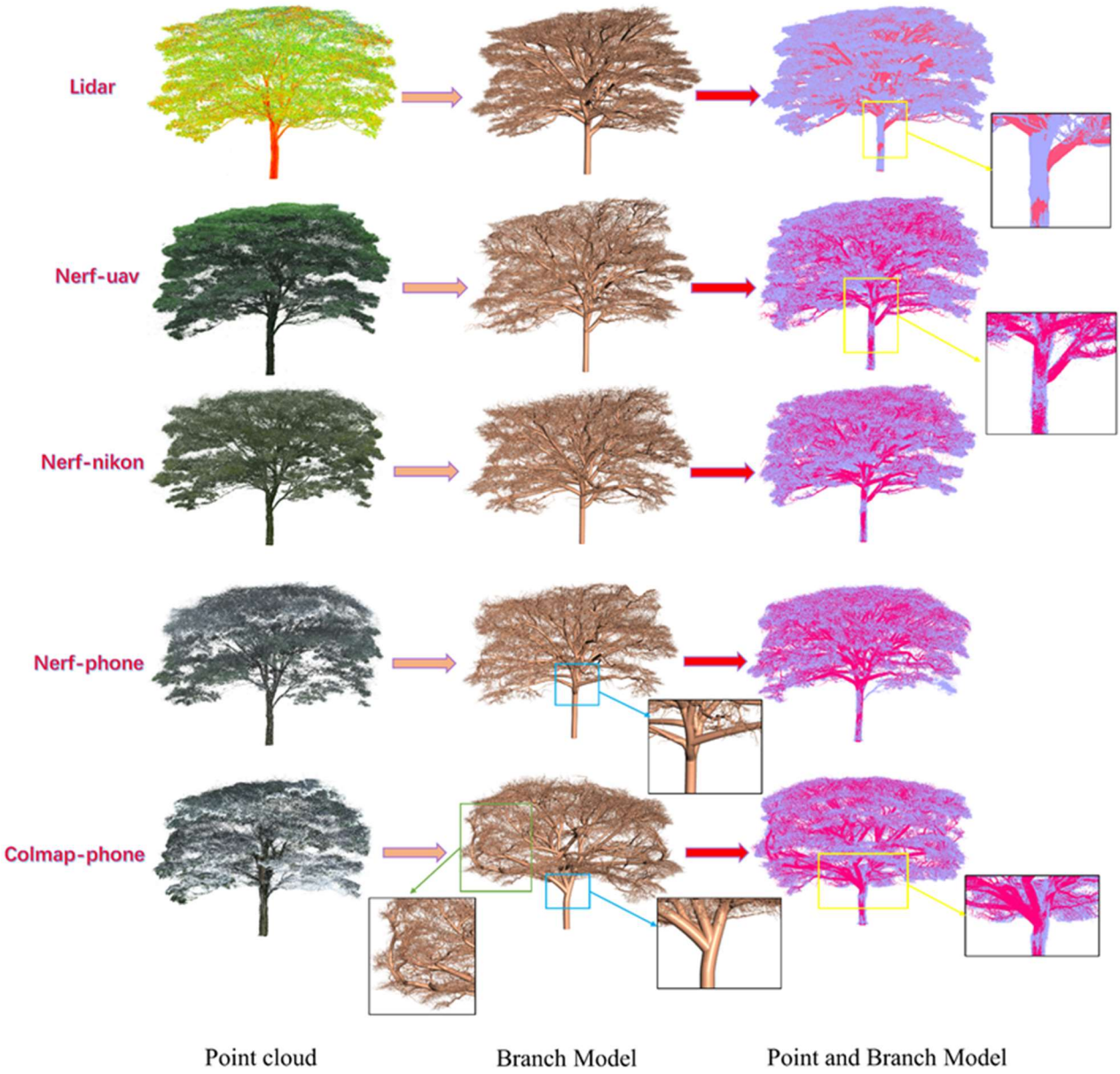


**Figure 7.** Tree_2 and its 3D branch model result reconstructed by AdTree. Left column is point cloud model, the LiDAR model is colored with intensity values and the NeRF and COLMAP models are colored in RGB; middle column is branch model; right column is the merged point and branch model, with the point cloud model color in purple and the branch model in red. The color boxes show the portions enlarged to highlight the details of the model.

## 4. Discussion

In this study, we compare two image-based 3D reconstruction methods for trees. One is based on the photogrammetry pipeline (SfM and MVS), while the other is based on neural radiance fields (using a set of images with known camera poses, usually obtained from the SfM procedure, to train and optimize a multi-layer perceptron network). Dense point clouds generated from these two methods were compared to the reference point cloud obtained from multi-station terrestrial laser scanning for reconstruction completeness and quality analysis.

The first observation is that 3D reconstruction will not always be successful. There is a requirement of a minimum number of images that must be met, but even for a large number of input images of trees and vegetative areas, the reconstruction can still fail. For Tree_2 in

our study, in the few cases when images acquired from the Nikon (107 images taken from the ground perspective) and drone (374 images taken both from the ground and in the air) were processed in COLMAP, partial reconstruction was achieved for the Nikon images (only 66 photos calibrated after SfM) and no dense reconstruction was achieved for the drone images. Concerned that the open source COLMAP might not be powerful enough, we also tested these two image datasets in the popular commercial photogrammetry software Metashape (Version 2.0.3) and Pix4dMapper (Version 4.7.5), obtaining similar failed or incomplete reconstruction results (Supplementary Material). In contrast, the NeRF method was more robust as it was able to handle these challenging situations well even in the face of limited input images or less-ideal image block configuration and still managed to produce visually appealing yet accurate outcomes using less time.

MVS utilizes the output of SfM to compute depth or normal information for every pixel in the image and then fuses the depth and normal maps of multiple images in 3D to produce dense point cloud of the scene. It is well recognized in computer vision that MVS faces challenges when dealing with thin, amorphous, self-similar or shiny objects, and trees with dense leaves or thin branches are these types of objects and they are especially difficult for point densification using the traditional photogrammetry method. NeRF can certainly play a role in filling the gap in this regard. Furthermore, MVS generally is a time-consuming process, and sometimes it takes days to process a dataset of a few thousand images. NeRF has the ability to greatly reduce the time required for reconstruction.

To ensure that the point cloud generated not only looks good but also represent the object or scene with high fidelity, we evaluated the metrics of cloud-to-cloud distances. For Tree_1, those two c2c distances (Mean $\pm$ STd) were very close: 0.031 $\pm$ 0.045 m for the COLMAP model vs. 0.037 $\pm$ 0.057 m for the NeRF model; for Tree_2, statistically, the NeRF-generated point cloud models all have smaller cloud-to-cloud distances, meaning they have closer spatial distribution to the reference TLS point cloud than those of the point cloud generated from MVS. This higher spatial proximity proves that the NeRF model can better represent the real scene in 3D.

The comparison of the tree structural parameter extraction illuminates the strengths and weaknesses of these two image-based tree reconstruction methods. We focus our discussion on the five parameters extracted in Table 4. NeRF-generated point clouds tend to be noisy, with large footprints or a lower resolution. In terms of tree height and DBH, current photogrammetry methods have advantages as they can offer results with a higher accuracy; for NeRF models of two trees, the best parameters extracted have a relative error of 2.4–2.9% in tree height and 7.8 to 12.3% error in DBH. However, in terms of canopy properties such as canopy width, canopy area and canopy volume, the NeRF method can provide better estimates than those derived from photogrammetry reconstruction models. And both types of point cloud can be fed into 3D tree modeling programs such as AdTree or TreeQSM [26] to create branch models and further can be used to estimate tree volume and carbon stock.

We used various cameras to take images or shoot videos. Video seems to have an edge in guaranteeing successful reconstruction results, as both of our studied trees were able to be reconstructed by the photogrammetric method using video frames collected by a smartphone camera. It is more efficient to record videos than take still photos, but the frames extracted from the video have lower resolution than normal images. For Tree_2, highly overlapped consecutive frames can be used to photogrammetrically reconstruct the tree, while images taken by other cameras (Nikon and UAV), even in similar manner (by shooting photos while walking or flying around the tree) but with less overlap, failed to accomplish that goal in COLMAP or other photogrammetric software programs. NeRF doesn't have this restrictive overlap requirement, and the best quality NeRF-generated point cloud was trained from UAV images, followed by Nikon images, with the worst being phone image frames. This demonstrates the importance of image resolution for NeRF reconstruction. Future field image collection may consider acquiring videos of at least 4K resolution. When the number of images to be processed is large, NeRF-based methods will

have the benefit of better efficiency. In addition, the weather and lighting conditions can also affect the reconstruction and rendering results.

Compared with traditional photogrammetry dense reconstruction methods such as MVS, NeRF is faster and has better reconstruction quality, especially for trees with dense canopies.

The result of NeRF rendering depends on the number of training epochs of the network, and it becomes stable after a certain threshold is reached. We investigated how many training iterations are needed in order for the network optimizer to converge, in part due to the fact that we need to pay an hourly charge for using the cloud-based GPU. We determined that when the number of training epochs reaches 10,000, the reconstruction results tend to stabilize. We may change the number of epochs and tweak other system parameters to further examine factors that might affect reconstruction outcomes.

Similar to the photogrammetry reconstructed model, models generated from NeRF methods are scale-less. We need to use markers or ground control points to manually bring them to the real physical world to make the reconstructed point cloud measurable. We noticed that the root mean squared error of the point cloud coarse and fine alignment was around 5 cm in CloudCompare, and we are pondering ways to reduce this alignment error and improve the point cloud accuracy.

Through this study, we have a better idea of the properties of NeRF-generated point cloud and have found answers to the questions raised in the Introduction; in the meantime, we envision many further research plans in this area: how will the accuracy of the input image pose parameters (three positions and two directions) affect the precision of the result; new approaches to reduce noise, improve the detail and resolution of the underlying scene representation; how to distill the NeRF into geometrically accurate meshes and point cloud; combination or fusion of NeRF, photogrammetry and laser scanning data; and large-scale forest scene applications for plot- and landscape-level forest inventory information collection and analysis.

In this study, we found that NeRF 3D reconstruction of single tree can achieve remarkable performance, and we believe there is great potential for future applications of NeRF to complex forest scenes. The technology is developing very fast, and we expect that more sophisticated and powerful tools like Gaussian Splatting [27] will be coming soon to make 3D forest scene reconstruction more affordable and accessible.

## 5. Conclusions

In this research, the NeRF technique is applied to 3D dense reconstruction of two trees with different canopy structures and compared with the traditional photogrammetric reconstruction technique. We tested the processing efficiency and capabilities of these two competing approaches using a series of images of trees acquired with different cameras and viewing angles. Quantitative and qualitative analyses were conducted to examine the visual appearance, reconstruction completeness, information content and utility of the NeRF- and photogrammetry-generated point clouds. Specifically, we looked into metrics including cloud-to-cloud distance, tree structural parameters and 3D models for comparison. The results show that

(1) The processing efficiency of the NeRF method is much higher than that of the traditional photogrammetric densification method of MVS, and it also has less stringent requirements for image overlap.
(2) For trees with sparse or little leaves, both methods can reconstruct accurate 3D tree models; for trees with dense foliage, the reconstruction quality of NeRF is better, especially in the tree crown area. NeRF models tend to be noisy though.
(3) The accuracy of the traditional photogrammetric method is still higher than that of the NeRF method in the extraction of single tree structural parameters in terms of tree height and DBH; NeRF models are likely to overestimate the tree height and underestimate DBH. However, canopy metrics (canopy width, height, area, volume

and so on) derived from the NeRF model are more accurate than those derived from the photogrammetric model.

(4) The method of image data acquisition, the quality of images (image resolution, quantity) and the photographing environment all have an impact on the accuracy and completeness of NeRF and photogrammetry reconstruction results. Further research is needed to determine the best practices.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs16060967/s1, Document S1: photogrammetry software comparison.docx.

**Author Contributions:** Conceptualization, H.H. and G.T.; methodology, H.H.; software, G.T.; validation, H.H. and G.T.; formal analysis, H.H. and G.T.; investigation, H.H. and G.T.; resources, H.H. and C.C.; data curation, H.H. and G.T.; writing—original draft preparation, H.H. and G.T.; writing—review and editing, H.H. and G.T.; visualization, G.T.; supervision, H.H.; project administration, H.H.; funding acquisition, C.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the University-Industry Cooperation Project in Fujian Province (Grant Number 2022N5008), International Cooperation Project of Fujian Province (Number 2022I0007) and Leading Talents of Scientific and Technological Innovation in Fujian Province, China.

**Data Availability Statement:** Data will be available upon reasonable request.

**Acknowledgments:** We would like to thank Cheng Li and Luyao Yang for their assistance in collecting and processing terrestrial laser scanning data of Tree_2 on campus.

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

## References

1. Kankare, V.; Joensuu, M.; Vauhkonen, J.; Holopainen, M.; Tanhuanpää, T.; Vastaranta, M.; Hyyppä, J.; Hyyppä, H.; Alho, P.; Rikala, J. Estimation of the timber quality of Scots pine with terrestrial laser scanning. *Forests* **2014**, *5*, 1879–1895. [CrossRef]
2. Wallace, L.; Lucieer, A.; Malenovský, Z.; Turner, D.; Vopěnka, P. Assessment of forest structure using two UAV techniques: A comparison of airborne laser scanning and structure from motion (SfM) point clouds. *Forests* **2016**, *7*, 62. [CrossRef]
3. Liang, X.; Kukko, A.; Balenović, I.; Saarinen, N.; Junttila, S.; Kankare, V.; Holopainen, M.; Mokroš, M.; Surový, P.; Kaartinen, H. Close-Range Remote Sensing of Forests: The state of the art, challenges, and opportunities for systems and data acquisitions. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 32–71. [CrossRef]
4. Iglhaut, J.; Cabo, C.; Puliti, S.; Piermattei, L.; O'Connor, J.; Rosette, J. Structure from motion photogrammetry in forestry: A review. *Curr. For. Rep.* **2019**, *5*, 155–168. [CrossRef]
5. Huang, H.; Zhang, H.; Chen, C.; Tang, L. Three-dimensional digitization of the arid land plant Haloxylon ammodendron using a consumer-grade camera. *Ecol. Evol.* **2018**, *8*, 5891–5899. [CrossRef] [PubMed]
6. Kükenbrink, D.; Marty, M.; Bösch, R.; Ginzler, C. Benchmarking laser scanning and terrestrial photogrammetry to extract forest inventory parameters in a complex temperate forest. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *113*, 102999. [CrossRef]
7. 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, 27–30 June 2016; pp. 4104–4113.
8. Xu, Z.; Shen, X.; Cao, L. Extraction of Forest Structural Parameters by the Comparison of Structure from Motion (SfM) and Backpack Laser Scanning (BLS) Point Clouds. *Remote Sens.* **2023**, *15*, 2144. [CrossRef]
9. Balestra, M.; Tonelli, E.; Vitali, A.; Urbinati, C.; Frontoni, E.; Pierdicca, R. Geomatic Data Fusion for 3D Tree Modeling: The Case Study of Monumental Chestnut Trees. *Remote Sens.* **2023**, *15*, 2197. [CrossRef]
10. Hinton, G.E.; Osindero, S.; Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [CrossRef] [PubMed]
11. 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. [CrossRef]
12. Yang, Z.; Chen, Y.; Wang, J.; Manivasagam, S.; Ma, W.-C.; Yang, A.J.; Urtasun, R. UniSim: A Neural Closed-Loop Sensor Simulator. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 1389–1399.
13. Corona-Figueroa, A.; Frawley, J.; Bond-Taylor, S.; Bethapudi, S.; Shum, H.P.; Willcocks, C.G. Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single X-ray. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, UK, 11–15 July 2022; pp. 3843–3848.

14. Chen, J.; Zhang, Y.; Kang, D.; Zhe, X.; Bao, L.; Jia, X.; Lu, H. Animatable neural radiance fields from monocular rgb videos. *arXiv* **2021**, arXiv:2106.13629.
15. Turki, H.; Ramanan, D.; Satyanarayanan, M. Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12922–12931.
16. Mazzacca, G.; Karami, A.; Rigon, S.; Farella, E.; Trybala, P.; Remondino, F. NERF for heritage 3D reconstruction. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *48*, 1051–1058. [CrossRef]
17. Condorelli, F.; Rinaudo, F.; Salvadore, F.; Tagliaventi, S. A comparison between 3D reconstruction using nerf neural networks and mvs algorithms on cultural heritage images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *43*, 565–570. [CrossRef]
18. Remondino, F.; Karami, A.; Yan, Z.; Mazzacca, G.; Rigon, S.; Qin, R. A critical analysis of nerf-based 3D reconstruction. *Remote Sens.* **2023**, *15*, 3585. [CrossRef]
19. Seitz, S.M.; Curless, B.; Diebel, J.; Scharstein, D.; Szeliski, R. A comparison and evaluation of multi-view stereo reconstruction algorithms. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; pp. 519–528.
20. Müller, T.; Evans, A.; Schied, C.; Keller, A. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.* **2022**, *41*, 102. [CrossRef]
21. 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 ICCV International Conference on Computer Vision, Virtual, 19–25 June 2021; pp. 5855–5864.
22. Tancik, M.; Weber, E.; Ng, E.; Li, R.; Yi, B.; Wang, T.; Kristoffersen, A.; Austin, J.; Salahi, K.; Ahuja, A. 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. 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.* **2021**, *40*, 237. [CrossRef]
24. Barron, J.T.; Mildenhall, B.; Verbin, D.; Srinivasan, P.P.; Hedman, P. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5470–5479.
25. Du, S.; Lindenbergh, R.; Ledoux, H.; Stoter, J.; Nan, L. AdTree: Accurate, detailed, and automatic modelling of laser-scanned trees. *Remote Sens.* **2019**, *11*, 2074. [CrossRef]
26. Raumonen, P.; Kaasalainen, M.; Åkerblom, M.; Kaasalainen, S.; Kaartinen, H.; Vastaranta, M.; Holopainen, M.; Disney, M.; Lewis, P. Fast automatic precision tree models from terrestrial laser scanner data. *Remote Sens.* **2013**, *5*, 491–520. [CrossRef]
27. Kerbl, B.; Kopanas, G.; Leimkühler, T.; Drettakis, G. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.* **2023**, *42*, 1–14. [CrossRef]