Viewing the Forest in 3-D: How Spherical Stereo Videos Enable Low-Cost Reconstruction of Forest Plots

Hristina Hristova, Arnadi Murtiyoso, Daniel Kükenbrink, Mauro Marty, Meinrad Abegg, Christoph Fischer, Verena C. Griess, and Nataliia Rehush

***Abstract*—Understanding and monitoring the surrounding environment increasingly rely on its 3-D representations. However, the often high costs of 3-D data equipment limit its wide usage, and low-cost solutions are in demand. Here, we propose a novel approach based on spherical stereo videos captured with a known baseline (distance between the cameras) for a low-cost and efficient 3-D point cloud reconstruction. In a forest environment, we evaluated 1) the influence of baseline length on point cloud quality and 2) the suitability of the generated point clouds for extracting primary forest attributes (tree position and diameter). Our results show that the proposed approach allows for feasible 3-D reconstruction of complex forest plots. The highest point cloud quality was achieved with a baseline of 60 cm. This setup enabled the correct detection of more than 65% of the trees within the forest plots, producing an average tree position error between 30 and 50 cm and clearly outperforming other setups. A multiscale model-to-model cloud comparison analysis showed signed distances between the generated point cloud and the reference data with zero mean and 1 m standard deviation. We demonstrate that the proposed approach can be a valuable low-cost solution for 3-D point cloud reconstruction, facilitating forest assessment and monitoring.**

***Index Terms*—3-D reconstruction, forest, low-cost, point cloud, spherical camera, stereo video, tree diameter, tree position.**

I. Introduction

GROUND-BASED 3-D capture technologies, such as terrestrial and mobile laser scanning (MLS) and close-range photogrammetry, have seen growing interest in forestry [1], [2]. They have shown an increased potential for producing highly detailed 3-D representations of forest sites and retrieving a wide range of forest parameters [3], [4], [5]. However, the operational use of such technologies is still limited due to high costs, substantial weight of devices, and significant data collection time. While MLS overcomes some of the limitations of terrestrial laser scanning (TLS) [6], [7], [8], it suffers from lower measurement precision [9].

In recent years, low-cost sensor solutions, such as low-end digital cameras, have demonstrated the ability to efficiently acquire data while substantially reducing the cost [10], [11]. These sensors are usually lightweight and readily accessible, with a robust, user-friendly interface for collecting reliable data, even for nonexperts. The benefits of low-cost sensors have extended to the photogrammetry field, where low-end cameras, such as mobile phone cameras and low-resolution spherical and fish-eye cameras, have been used to replace high-end cameras for reconstructing scenes from multiple stereo images [12], [13], [14].

In forestry, photogrammetric methods are used to create 3-D representations of forest sites by capturing multiple overlapping forest images with middle-range to high-end cameras in a stop-and-go or mobile manner. This technology has gained significant attention in recent times [15], [16], [17], [18]. Forest point clouds have also been generated using enhanced smartphone photographs as an alternative to images from high-end cameras [19]. Furthermore, the time-lapse function of the GoPro sensor (a fish-eye lens) has been exploited for collecting wide-lens forest content to build point clouds [1]. Furthermore, the potential of using forest images for monocular depth estimation using deep learning has recently been investigated [20]. A review of the deep learning methods and their prospects for various forest conditions and different types of remote sensing data has been detailed in [21]. Videogrammetry, as a subset of photogrammetry, encompasses the process of 3-D reconstruction from video sequences [22]. Video capture is a faster and more robust way to acquire data than capturing stereo images [23], although often at the expense of image quality. The overlapping video frames extracted from a video sequence can then successfully be processed using conventional photogrammetric principles to generate point clouds. Videogrammetry has mainly been used in the context of indoor scenes and cultural heritage [13], [24], [25], [26], [27], with a more recent focus on low-end spherical videos [24], [25], [28]. However, compared with point clouds,

Received 30 April 2024; revised 4 August 2024; accepted 7 September 2024. Date of publication 18 September 2024; date of current version 7 October 2024. This work was supported by the Swiss National Forest Inventory through the scientific project "Potential of 360° images for applications of the Swiss NFI." *(Corresponding author: Hristina Hristova.)*

Hristina Hristova, Daniel Kükenbrink, Mauro Marty, Meinrad Abegg, Christoph Fischer, and Nataliia Rehush are with the Swiss Federal Institute for Forest, Snow, and Landscape Research WSL, CH-8903 Birmensdorf, Switzerland (e-mail: hristina.hristova@wsl.ch; daniel.kuekenbrink@wsl.ch; mauro.marty@wsl.ch; meinrad.abegg@wsl.ch; christoph.fischer@wsl.ch; nataliia.rehush@wsl.ch).

Arnadi Murtiyoso and Verena C. Griess are with the Institute of Terrestrial Ecosystems, ETH Zurich, CH-8092 Zurich, Switzerland (e-mail: arnadi.murtiyoso@usys.ethz.ch; verena.griess@usys.ethz.ch).

Digital Object Identifier 10.1109/JSTARS.2024.3462999

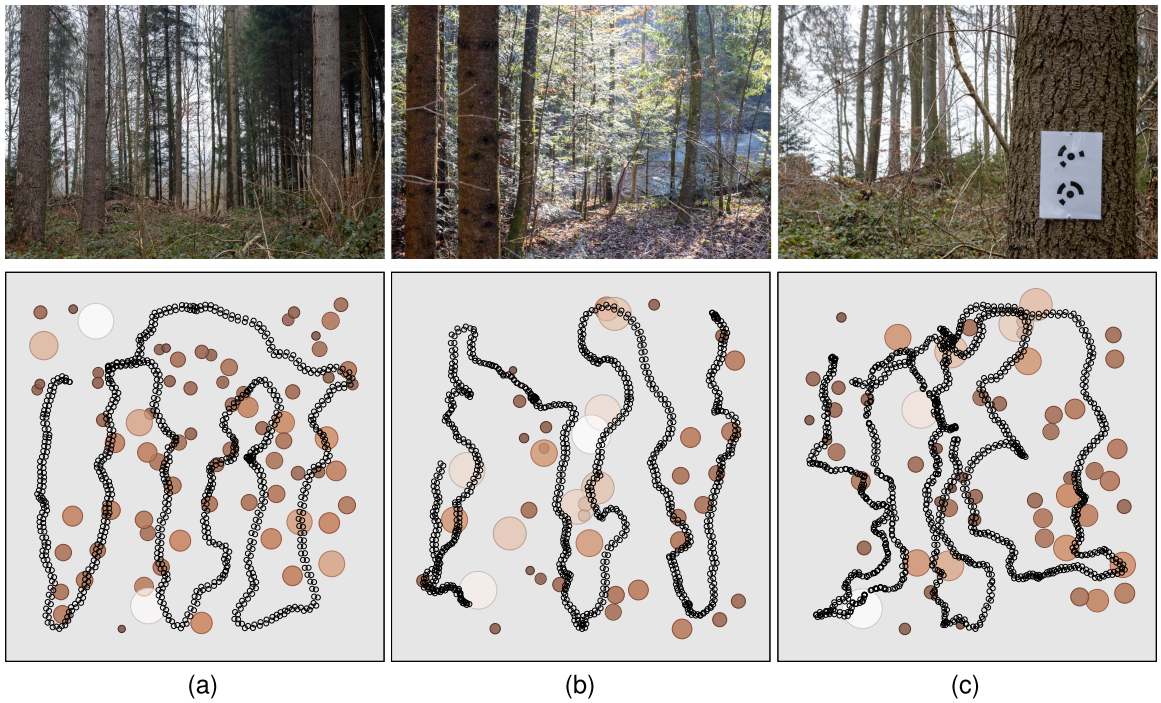

Fig. 1. Forest study area and acquisition patterns. Three plots in temperate mixed forest were selected (a)–(c), representing both sparse and dense forest conditions. An example of the visually coded targets distributed throughout the study area is shown in image (c). The second row contains the tree distributions per plot. The trees are visualized by circles, whose diameter corresponds to the DBH on a 1:170 scale. In black, we illustrate the acquisition patterns per plot and show representative camera positions only. (a) Plot 1. (b) Plot 2. (c) Plot 3.

which were built with classical photogrammetry or acquired using TLS technology, videogrammetric point clouds have shown signs of noise due to the lower image resolution of the video frames [29]. This limitation has prompted research into the effects of blur filters, video frame rate, and video resolution on the quality of a 3-D videogrammetric reconstruction [25], [29], [30].

Unlike classical photogrammetry, videogrammetric approaches are rare in forestry. Low-end spherical cameras have shown great potential for video acquisition in forest environments, as they provide a complete view of a forest scene [31], [32]. Murtiyoso et al. [31] used a Ricoh Theta Z1 dual-lens spherical camera to record monocular videos in a forest environment, followed by a 3-D reconstruction of the geometry of habitat trees and mapping of the forest understory. The authors employed videogrammetric principles to produce dense point clouds, which they then scaled using evenly distributed visually coded targets on the forest plot. Setting up such visually coded targets can significantly increase the acquisition time and add complexity to the data collection. Furthermore, the quality of the generated point cloud may depend on the expertise of the person placing the targets in the field and how well the targets are detected by the photogrammetric software [33]. Uneven or ill-designed networks of these targets may also generate 3-D model distortions [34], [35]. Alternatively, a stereo setup with a known baseline (distance between two camera sensors) fixed through the data acquisition process has been used for absolute scaling in conventional pinhole cameras [36]. However, Dai et al. [36] concluded that there was a necessary tradeoff between baseline length and 3-D reconstruction precision. Indeed, based on the epipolar model [37], smaller baselines theoretically present more risk to geometric stability and, thus, also quality. To the best of the authors' knowledge, no studies have been performed to evaluate the influence of baseline length on the quality of point cloud reconstruction from stereo videos collected in a forest environment.

In this study, we propose a novel videogrammetric approach based on spherical stereo videos for low-cost 3-D reconstruction of forest sites. Our method involves a stereo setup with a known baseline that eliminates the need for visually coded targets, making point cloud building and scaling more efficient and straightforward. The key objective of this study is to evaluate the potential of the stereo-video approach for the 3-D reconstruction of entire forest plots. For this, we 1) evaluated the influence of the baseline in a stereo-video setup on the quality of the generated point clouds, 2) compared the performance of the stereo-video and monocular-video approaches, and 3) evaluated the feasibility of the videogrammetric point clouds to extract primary forest attributes, such as tree position and diameter at breast height (DBH). We conducted our experiments on three forest plots using stereo equipment consisting of two spherical cameras. We tested three baselines (i.e., 20, 40, and 60 cm) and generated point clouds based on the captured video content. In addition, we reconstructed point clouds from monocular videos with visually coded targets distributed within the forest plots. Finally, we assessed the quality of the videogrammetric point clouds with the Terrestrial Laser Scanning (TLS) data used as reference.

Fig. 2. Stereo video system used in this study targeted two dual-lens fish-eye Ricoh Theta Z1 cameras with a known baseline (distance between the camera sensors). To foster comparability, we used three spherical cameras arranged on a hand-held gimbal to simultaneously capture stereo videos with three baselines, namely, 20, 40, and 60 cm.

II. Data and Methods

A. Study Area

The study area is located in a temperate mixed forest in the canton of Zurich, Switzerland (47° 21'40"N, 8° 27'10"E). The main tree species in the study area is European beech (*Fagus sylvatica* L.), followed by silver fir (*Abies alba* Mill.) and Norway spruce [*Picea abies* (L.) H. Karst.]. We conducted our experiments on three forest plots, each with a size of 50 m × 50 m (see Fig. 1). Plot 1 comprised a relatively homogeneous forest with trees of similar sizes, while Plots 2 and 3 represented complex heterogeneous forests with various tree sizes [see Fig. 1(b) and (c)]. More details about the forest plots used in this study can be found in [1].

B. Data Acquisition

We acquired the data using 1) a stereo setup of digital cameras, 2) a single digital camera, and 3) a terrestrial laser scanner. We acquired the data in February 2023 under leaf-off conditions to minimize occlusion effects by avoiding dense foliage.

1) Stereo Video Capture: Stereo-video capture involved three spherical cameras, allowing for the simultaneous recording of stereo videos with three baselines: 20, 40, and 60 cm. We used Ricoh Theta Z1 spherical sensors (Ricoh Company, Ltd., Tokyo, Japan), each featuring two fish-eye lenses, and recorded videos with 4 K resolution, 30 frames per second (fps), and 7.3 mm focal length (35 mm equivalent). The cameras were attached to a hand-held gimbal (Neewer, Shenzhen, China) using a specially designed camera mount (see Fig. 2). The first two cameras (left to right) were placed 40 cm apart, while the second and third cameras were placed 20 cm apart. With this setup, we eliminated acquisition bias caused by variations in path and shaking effects during separate acquisitions.

During data acquisition, the operator walked through the forest plot at moderate speed, which resulted in completing data acquisition in around 8–10 min per plot. The operator's walking speed was not always constant due to the heterogeneous structure of the forest floor (i.e., the presence of dense vegetation and lying deadwood). We recorded the videos following an acquisition grid pattern [25], as shown in Fig. 1. To avoid any on-the-fly stitching artifacts, we acquired all videos in raw format. We then merged the raw contents of the fish-eye lenses offline using the Ricoh Theta Z1 stitching application. To synchronize the stereo videos, we searched for matching sound patterns utilizing the DaVinciResolve video editing software [38]. We ensured that each pair of synchronized videos started with the same sound pattern. Moreover, we registered the time using a stopwatch during the video acquisition, showing time up to the milliseconds. We visually checked the synchronization by inspecting the video frame pairs and comparing them according to the stopwatch time values. Here, we further denote the stereo-video methods, utilizing baselines of 20, 40, and 60 cm, as `stereo-20`, `stereo-40`, and `stereo-60`, respectively.

2) Monocular Video Capture: Along with the stereo-video acquisition, we captured three monocular videos (one for each spherical camera in our acquisition equipment). Reusing the same videos (but in a monocular-video manner) allowed us to enhance comparability by reducing bias from separate acquisitions. To give an absolute scale to the 3-D point cloud, we distributed a set of 24 visually coded targets evenly throughout the plots, following conventional 3-D aerotriangulation and

absolute orientation requirements [34]. Each target comprised two circular 12-bit coded targets situated a known distance apart, as shown in Fig. 1(c). Here, we set this distance to 13.4 cm. It served as a scaling factor when building point clouds using the `monocular-video` approach, as introduced and elaborated in [31].

3) TLS Acquisition: For reference purposes, we used a TLS point cloud acquired with a Riegl VZ400i (RIEGL Laser Measurement Systems GmbH, Horn, Austria). For the acquisition, we chose a regular grid pattern with a 10 m distance between consecutive scan positions, following recommendations in [39]. We performed a vertical and 90°-tilted scan for each scan position to cover the full sphere, as the vertical field of view of the Riegl VZ400i scanner only covers 100°. We set the scanning resolution to 0.04°, resulting in a scan time of 45 s per scan. All scan positions were successfully automatically coregistered using the built-in scan registration capabilities of the Riegl VZ400i TLS instrument. We processed all scans using the Riegl RiScan Pro processing software (Riegl, v2.16.1). After applying Riegl's multistation adjustment procedure, we achieved a final average standard deviation of the registered point clouds of 0.0013 m, with values for individual scans ranging from 0.0006 to 0.0037 m (calculated from 277 748 filtered patches).

C. Point Cloud Generation From Stereo Videos

We exploited the stereo-video content captured during data acquisition to reconstruct point clouds of the study area using the spherical photogrammetric pipeline in the the AgiSoft Metashape software [40]. From the stitched versions of the stereo videos, we extracted video frames with a frequency of 5 fps and imported them into AgiSoft Metashape software. We studied also higher frame frequencies, such as 15 and 30 fps. However, we found no significant increase in point cloud quality, where the computational time doubled.

The two key steps of classical point cloud generation in the AgiSoft Metashape software are 1) image (frame) orientation and 2) dense depth map estimation. We added an extra step before the image orientation step to introduce scale into the AgiSoft Metashape project. For this, we defined a scale factor between each pair of stereo frames, the value of which was equal to the baseline with which the stereo videos were captured by creating so-called scale bars. We created these scale bars and assigned the corresponding baseline value to each scale bar using a Python script. Next, we performed a frame orientation with "medium" alignment settings. The cameras were auto-calibrated using the spherical calibration mode in the AgiSoft Metashape software. We provided no external camera intrinsics for the calibration. After successful orientation, we built a dense point cloud with "medium" quality and "mild" filtering parameters. A "medium" quality implies the employment of stereo multiview dense matching on subsampled pixels of the original image, amounting to 1/16 of the original pixel resolution. At the same time, the filtering parameters refer to the specific depth map filtering strategies used by the software. While the algorithm behind this filtering is not publicly available, a type of epipolar constraint may be involved [41].

Algorithm 1: Algorithm for Tree Position and DBH Extraction From the Normalized Point Cloud P. The Main Functions (i), where $i \in [1, 5]$, Correspond to the Steps in our Pipeline.

procedure FILTERBYVERTICALITY(P, r, t)
  P_v = VERTICALITY(P, r)
  P_f = ApplyThreshold(P_v, t)
  return P_f
procedure DENOISE(P, n, s, l_o, p_{min})
  P_{sor} = SOR(P, n, s)
  $\{P_{ccl}\}_{i=1}^{N}$ = ComputeCCL(P_{sor}, l_o, p_{min})
  P_d = Merge($\{P_{ccl}\}$)
  return P_d
procedure EXTRACTTREESTEMS(P, r_1, r_2, t_1, t_2)
  P_f = FilterByVERTICALITY(P, r_1, t_1)
  P_{stems} = FilterByVERTICALITY(P_f, r_2, t_2)
  return pc_{stems}
procedure FILTERTREESTEMS(P)
  P_d = Denoise(P, 20, 0.1, 10, 1000) ▷(1)
  P_{stems} = ExtractTreeStems(P_d, 0.3, 0.1, 0.65, 0.75) ▷(2)
procedure EXTRACTTREEPOSITIONS(P_{stems})
  $\{stem_i\}_{i=0}^{M}$ = ComputeCCL(P_{stems}, 10, 2000) ▷(3)
  for each $stem_i$ **do**
    Pos_i = ProjectToXY($stem_i$, H, T) ▷(4)
  procedure ESTIMATEDBH($\{pos_i\}_{i=0}^{M}$) ▷(5)
  for each pos_i
    $hull_i$ = ConvexHull(pos_i)
    D_i = RANSAC(pos_i, $hull_i$)

D. Point Cloud Generation From a Monocular Video

We used the AgiSoft Metashape software to create point clouds from the three monocular videos collected with our equipment. Generating a point cloud from a monocular video sequence relied on visually coded targets to provide the scale. The AgiSoft Metashape software detected these targets after orienting the video frames. We then created scale bars between each pair of detected targets before recomputing the orientation in a block bundle adjustment process and building a dense point cloud via multiview stereo considering the monocular-video sequential frames. We used the "medium" setting to auto-calibrate and orient the video frames. We built the dense point clouds with "medium" quality and "mild" filtering parameters, similar to the stereo-camera approach. For each forest plot, we generated three monocular-video point clouds. For our analysis, we selected the highest-quality point cloud among the three. All point clouds were built on a Mac machine with a dedicated Radeon Pro 580 GPU with 8 GB of memory.

E. Point Cloud Preprocessing

The preprocessing steps involved 1) coregistration of all videogrammetric and TLS point clouds and 2) point cloud normalization. These steps are required to evaluate properly the quality of the generated videogrammetric point clouds. We

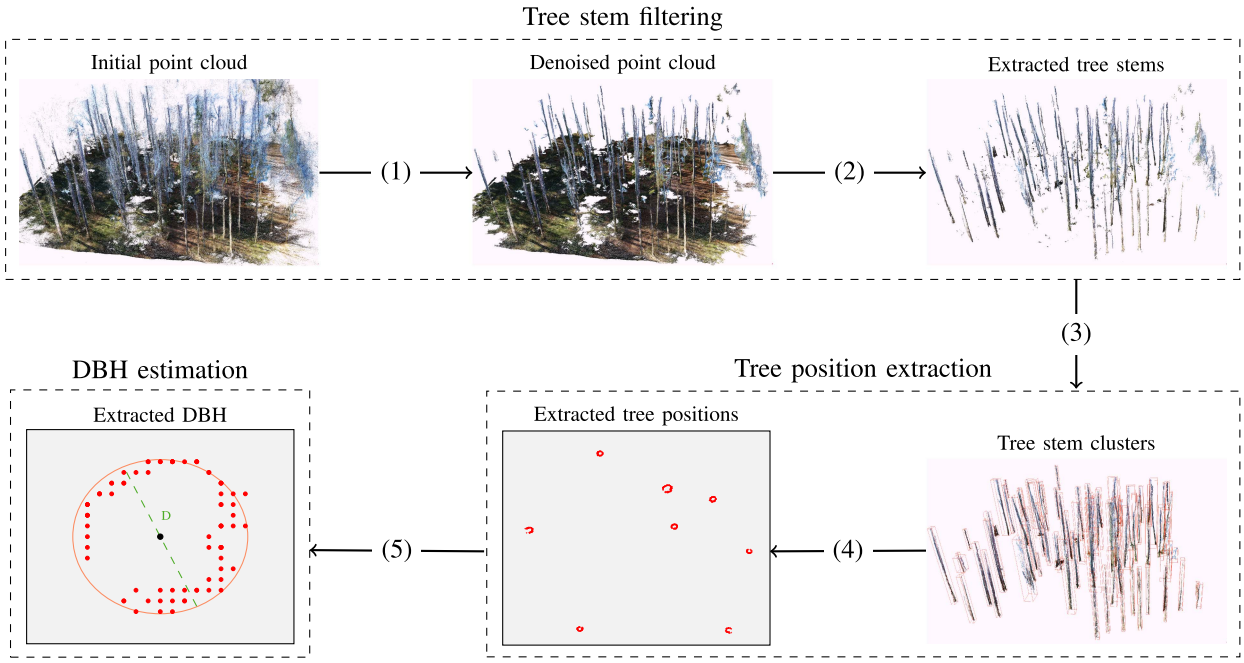

Fig. 3. Tree position and DBH extraction pipeline. The main steps, denoted as (i), where $i \in [1, 5]$, correspond to functions in Algorithm 1. The visualized extracted tree positions showcase a snippet of the entire set of extracted tree positions on the forest plot.

completed the point cloud coregistration in two stages: 1) we manually performed a rough coregistration of the two given point clouds via common points, and 2) we used the iterative closest point (ICP) fine registration method implemented in the CloudCompare software [42] to refine the coregistration further.

Based on our goal to determine and compare the position and DBH of trees within the forest plots, it was necessary to normalize both the TLS data and the videogrammetric point clouds, removing the influence of the terrain. We performed the normalization of all point clouds according to the ground extracted from the TLS point clouds, as we used TLS data as reference data in our study. To classify the ground points in the TLS point clouds, we used the cloth simulation filter (CSF) [43] implemented in the CloudCompare software.

F. Extraction of Forest Attributes

To evaluate the effectiveness of our videogrammetric approach for forest applications, we assessed its potential to derive primary forest attributes, such as tree position and DBH, from the generated videogrammetric point clouds.

1) Tree Position and Diameter at Breast Height (DBH) Extraction: We extracted tree positions and DBH from normalized videogrammetric and TLS point clouds as described in Algorithm 1 and illustrated in the pipeline in Fig. 3. The tree position extraction consisted of two main steps: stem filtering and tree position derivation.

First, we applied gentle noise filtering to the normalized point cloud using the statistical outlier removal (SOR) with a neighbor count n of 20 and a standard deviation threshold t of 0.1 m. Next, we used the connected components labeling (CCL) algorithm [44] to cluster the point cloud into connected segments. These segments contained a minimum number of points per component p_{min} equal to 1000. The result of merging the connected segments is shown in Fig. 3 as "denoised point cloud." In the next step, for the denoised point cloud, we computed the `verticality` feature with a neighborhood radius r of 0.3 m and kept only points with a value above a threshold t of 0.65. This filtering operation was repeated one more time with a radius of $r = 0.1$ m, retaining only points with a `verticality` value above 0.75. These operations preserved tree stems and eliminated most ground points, leaves, and small branches (see Fig. 3).

We clustered the point cloud containing the extracted tree stems using the CCL algorithm at an octree level l_o of 10 with $p_{\text{min}} = 2000$, ensuring that each cluster contained an individual tree stem. For each cluster, we computed a cross-section at a height H of 1.3 m with a thickness T of 20 cm (10 cm above and below 1.3 m) and projected it onto the XY plane. Lastly, we associated the extracted tree positions with the center of each projection. We derived all the function parameters discussed in this section and described in Algorithm 1 empirically, maximizing the number of extracted trees. We applied the same function parameters to all videogrammetric and TLS point clouds.

To estimate tree DBH, we applied a modified version of the random sample consensus (RANSAC) algorithm [45]. The RANSAC algorithm determines the best-fitting circle for the points within each projection. We constrained our DBH estimation by half the size of the convex hull calculated for each projection, which helped us achieve an optimal circular fit.

2) Reference Tree Matching: To enhance point cloud comparability, we matched reference trees from the TLS data to trees from the generated videogrammetric point clouds. Only trees closer than 2 m from the reference positions were considered a

match. When a reference tree was matched to multiple trees from a videogrammetric point cloud, we used a recursive function to identify the closest tree. This ensured that each reference tree was only paired with one tree from the generated point cloud and that the paired tree was the nearest match.

G. Point Cloud Quality Assessment

We assessed the quality of the generated videogrammetric point clouds by computing camera orientation errors during point cloud reconstruction and calculating the multiscale model-to-model cloud comparison (M3C2) error. Furthermore, we evaluated the potential of the generated videogrammetric point clouds for forest applications by computing tree detection rate, absolute tree position error, and relative DBH error. We computed the evaluation metrics relative to the reference TLS data. To assess the statistical significance of the differences in the tree position distributions, as well as DBH and M3C2 distributions, calculated for each videogrammetric approach, we used the nonparametric Mann–Whitney independence test.

1) Camera Orientation Error: To assess 3-D reconstruction quality, we computed camera orientation errors for each videogrammetric approach. These errors are derived from the camera positions predicted with the Agisoft Metashape software. The errors represent how much the distance between corresponding stereo positions deviates from the scale bar value set in the project.

2) Multiscale Model-to-Model Cloud Comparison: We used the M3C2 metric [46] to measure cloud-to-cloud distance by detecting local distance changes in reference core points extracted from the TLS data. We computed the Multiscale Model-to-Model Cloud Comparison (M3C2) metric for each core point based on its distance to neighboring points in the videogrammetric point clouds. We ignored the core points with no neighboring points. To assess the quality of different parts of the point cloud, we computed the M3C2 metric separately for tree stems and ground points using the M3C2 implementation in the CloudCompare software. To extract tree stems, we used the *FilterTreeStems(·)* function in Algorithm 1. To prevent any loss of information, we applied this function to the nonnormalized point clouds after extracting ground points using the CSF algorithm implemented in the CloudCompare software [43].

3) Tree Detection Rate: We defined the tree detection rate r_d as the percentage of trees successfully detected in the videogrammetric point cloud in relation to the number of trees identified in the reference TLS data

$$r_d = 100 * \frac{m}{N} \tag{1}$$

where m is the number of trees in the videogrammetric point cloud matched to reference TLS trees, and N is the number of extracted trees from the TLS point cloud.

4) Absolute Tree Position Error: We calculated the absolute position errors for each tree in a videogrammetric point cloud to determine how close they were to the TLS data using the following formula:

$$e_{\text{pos}} = |c_v - c_{\text{tls}}| \tag{2}$$

TABLE I
CAMERA ORIENTATION ERRORS DERIVED FROM THE CAMERA POSITIONS ESTIMATED WITH THE AGISOFT METASHAPE SOFTWARE

	Camera orientation error [mm]		
Approach	Plot 1	Plot 2	Plot 3
stereo-20	1.265	2.496	2.105
stereo-40	0.906	1.833	0.966
stereo-60	0.628	0.770	1.080
single-camera	4.261	5.636	28.025

where c_v and c_{tls} are the extracted positions of the trees in a given videogrammetric point cloud and their counterparts in the corresponding TLS point cloud, respectively.

5) Relative DBH Error: To assess the proximity of an estimated DBH to the reference DBH, we computed the relative errors e_{DBH} as follows:

$$e_{\text{DBH}} = 100 * \frac{|D_v - D_{\text{tls}}|}{D_{\text{tls}}} \tag{3}$$

where D_v and D_{tls} are the extracted tree DBH of matched trees in a given videogrammetric point cloud and the corresponding TLS point cloud, respectively.

III. RESULTS

A. 3-D Reconstruction Quality

We successfully generated point clouds from all acquired stereo-video and monocular-video sequences. As shown in Table I, the `stereo-60` approach achieved the smallest camera orientation errors. In contrast, the largest camera orientation errors were computed for the `monocular-video` approach, with values exceeding the sought-after accuracy error of 1mm by four to five times for Plot 1 and Plot 2 and by 28 times for Plot 3. These results indicate that the camera orientation in the AgiSoft Metashape software is more successful with the stereo-video approach than the `monocular-video` approach and that the orientation process benefits from the extra stereo information and the known baseline between the stereo videos.

B. Tree Detection Rate

The `stereo-60` approach produced the highest tree detection rate of 71% for Plot 1, 68% for Plot 2, and 65% for Plot 3 [see Fig. 4(a)]. For Plot 1 and Plot 2, both the `stereo-40` and `monocular-video` approaches had similar performance. However, for Plot 3, the `monocular-video` approach only reconstructed 22% of the reference trees.

Tree size impacted the detection of trees in the videogrammetric point clouds [see Fig. 4(b)]. To study this impact, we combined all trees from Plots 1–3 and stratified them into five categories according to their DBH, estimated from the corresponding reference TLS data. The lowest detection rate (16% for `stereo-40` and 33% for `stereo-20`, `stereo-60`, and `monocular-video`) was computed for trees with a DBH of less than 10 cm. In contrast, the detection rates improved for

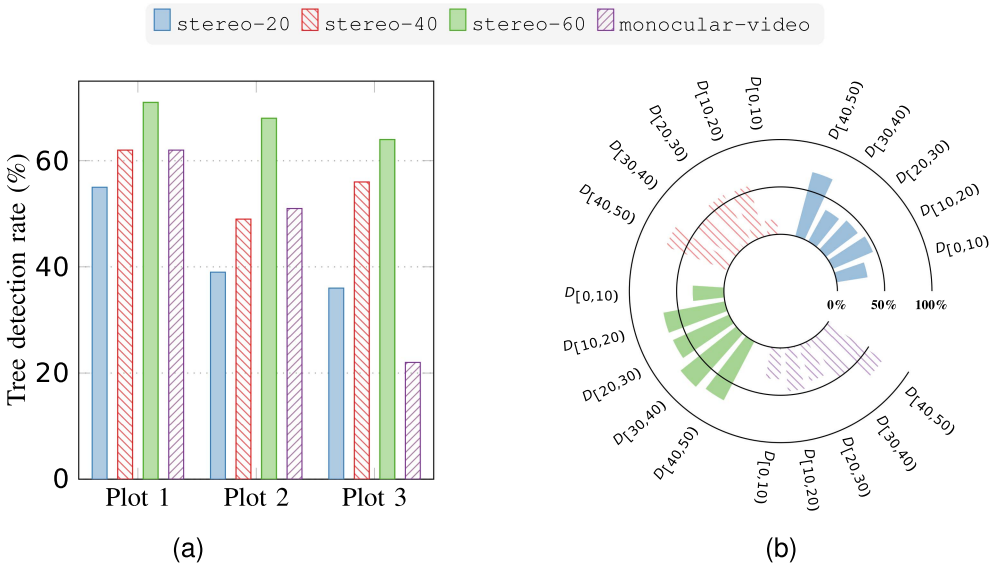

Fig. 4. Tree detection rate for all videogrammetric approaches (a) per forest plot and (b) per tree diameter category computed relative to the reference TLS data. We considered five tree diameter categories $D_{[u,l)}$, where D is the estimated DBH and $D \in [u, l)$.

trees with a DBH above 10 cm, with the best overall rate being for trees with a DBH above 40 cm. In general, the `stereo-60` approach achieved the highest detection rates for all tree sizes.

C. Tree Position Extraction

The tree positions for all the forest plots extracted at a height of 1.3 m from the videogrammetric point clouds and the corresponding TLS point clouds are visualized in Fig. 5(a). Moreover, Fig. 5(b) shows the tree position error distribution (in m) for all videogrammetric approaches. The `stereo-60` approach yielded the lowest median tree position error over all three forest plots, with 30 cm for Plot 1 and less than 50 cm for Plots 2 and 3. The second-best result was obtained for the `monocular-video` method, followed by the `stereo-40` and `stereo-20` approaches.

The Mann–Whitney independence test indicated that the difference in the tree position error distributions between the `stereo-60` and `monocular-video` approaches was statistically significant, with a p-value of less than 0.0001 for Plot 1 and less than 0.05 for the more complex Plots 2 and 3. Moreover, the tree position error distributions for the `stereo-20`, `stereo-40`, and `stereo-60` approaches were statistically different, with a single exception for Plot 2. The tree position errors of the `stereo-20` and `stereo-40` approaches for Plot 2 come from the same population, i.e., the two approaches are interchangeable regarding tree position precision for this specific plot.

The analysis of the tree position errors reveals that the proposed stereo-video approach with a baseline of 60 cm outperformed the `monocular-video` approach. On average, the tree positions computed with the `stereo-60` were 20–50 cm closer to the reference tree positions than those computed with the `monocular-video` approach [see Fig. 5(b)]. Finally, according to the Mann–Whitney test, for Plot 1 and Plot 3, the `monocular-video` approach was statistically indistinguishable from the `stereo-40` approach.

D. DBH Extraction

For all plots, the `stereo-60` and `stereo-40` approaches performed the best with a median DBH error of less than 25%. For Plots 1 and 2, all videogrammetric approaches performed similarly concerning the precision of DBH extraction (see Fig. 6). The Mann–Whitney independence test showed that for these two plots, the relative DBH error distributions of all videogrammetric approaches were statistically indistinguishable. Plot 3, however, proved challenging for the `monocular-video` approach, with an increase in the median DBH error to 38%.

We evaluated the influence of tree size on DBH extraction precision based on the `stereo-60` approach [see Fig. 6(d)]. Over 80% of trees with a DBH greater than 30cm had a relative DBH error (e_{DBH}) of less than 20%, while more than 40% of them had a e_{DBH} of less than 10%. None of these trees showed huge errors (over 50%). Conversely, 11% to 17% of the trees whose DBH was between 10 and 30 cm contributed to DBH errors greater than 50%. Furthermore, over 70% of the trees with a DBH below 10 cm had e_{DBH} greater than 20%, which signifies that the DBH estimation from the videogrammetric point clouds was inefficient for thin trees.

E. Cloud-to-Cloud Comparison

We assessed the quality of the point cloud reconstruction using the M3C2 distance computed for all videogrammetric point clouds relative to the reference TLS data. The metric was calculated separately for tree stems and ground points, as depicted in Section II-G2.

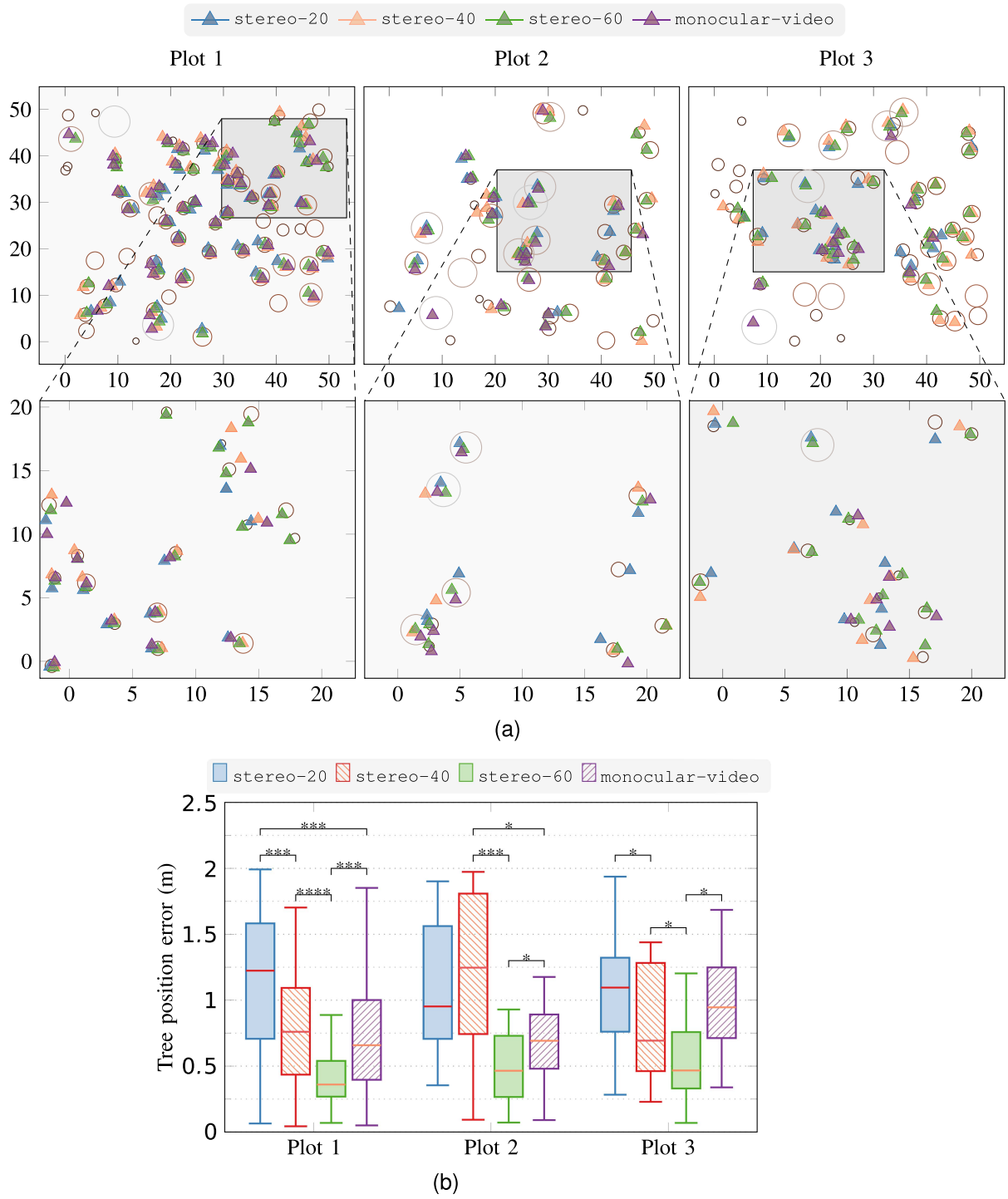

Fig. 5. (a) Tree position visualizations. Circles represent tree positions from the TLS, scaled according to the estimated DBH. Triangles represent the tree positions extracted from videogrammetric point clouds. The indicated scale is in meters. (b) Tree position error distributions for the three stereo-video approaches and the `monocular-video` approach compared with TLS. The statistical significance of the distribution difference is indicated with (1) * for p-value < 0.05, (2) *** for p-value < 0.001, and (3) **** for p-value < 0.0001.

1) Tree Stems: For parts of all videogrammetric point clouds associated with tree stems, the M3C2 errors were normally distributed with approximately zero means (see Fig. 7). The largest standard deviation of 1.96 m was obtained for the `stereo-20` approach. This approach was prone to large errors, as observed in the color maps (first two columns in Fig. 7). In contrast, most of the M3C2 errors computed for the `stereo-60` approach were distributed around zero, with the smallest standard deviation (equal to 1 m) among all four approaches. In addition, the `stereo-60` method reconstructed more tree stems than the other videogrammetric approaches (fewer gray areas) and had significantly lower reconstruction errors. On the other hand, the

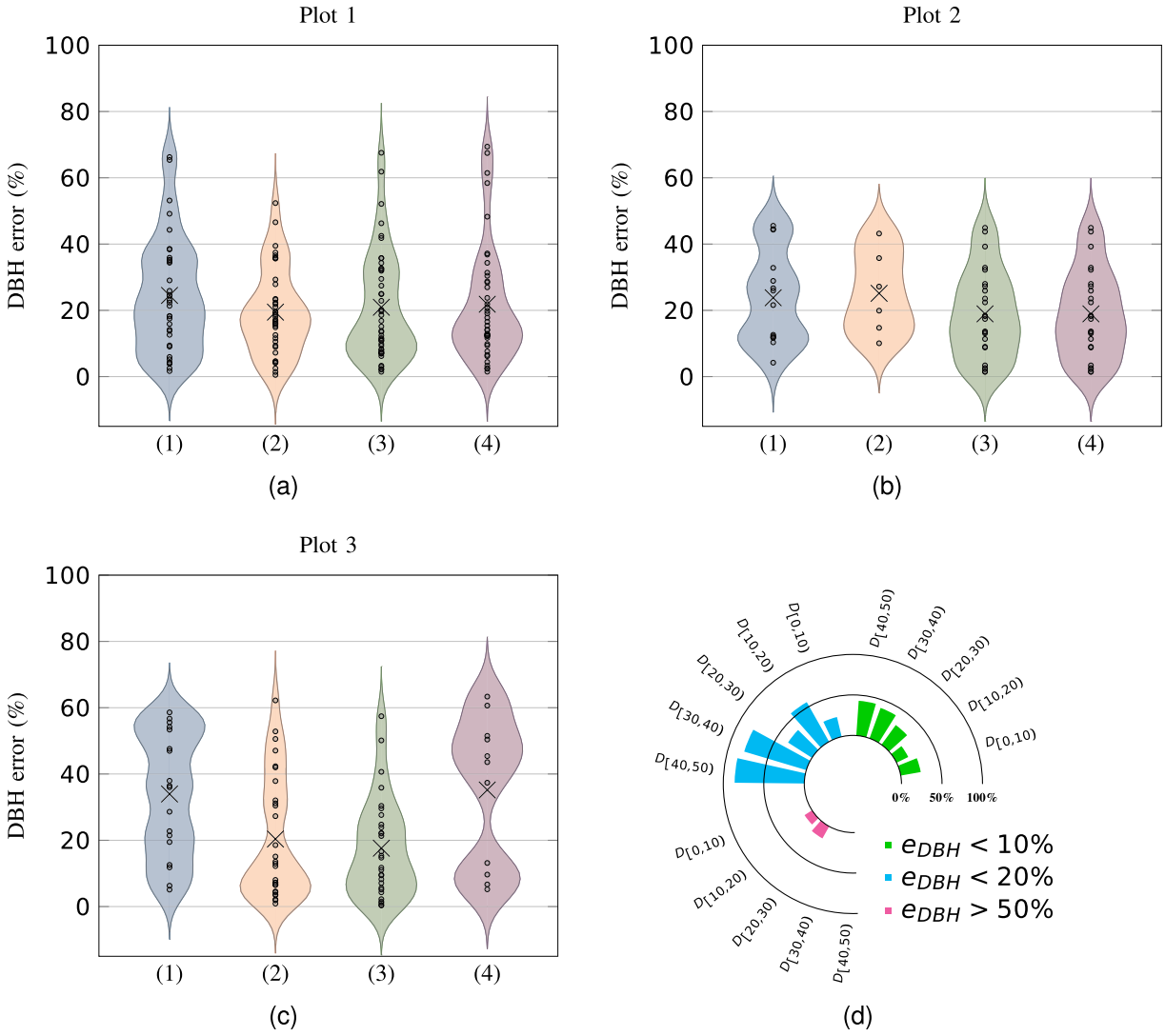

Fig. 6. (a)–(c) Distributions of relative tree DBH errors for the (1) `stereo-20`, (2) `stereo-40`, (3) `stereo-60`, and (4) `monocular-video` approaches. The black circles show diameter error data, which are always positive. The cross in each violin plot indicates the median value of the distribution. The negative curve values are a result of curve smoothing. (d) Percentage of trees per diameter category corresponding to various DBH errors e_{DBH} computed for the `stereo-60` approach.

`stereo-40` and `monocular-video` approaches had statistically similar distribution densities with identical standard deviations. Finally, all videogrammetric approaches reconstructed the tree stems to a certain height. The tree crowns were not recovered.

2) Ground: The M3C2 distance distributions for the ground points of the videogrammetric point clouds for Plot 1 are visualized in Fig. 8. The Mann–Whitney independence test confirmed that the M3C2 distributions computed for the ground points of all videogrammetric approaches were statistically interchangeable. This finding means that the `stereo-20`, `stereo-40`, `stereo-60`, and `monocular-video` methods reconstructed the forest ground similarly.

F. Time Assessment

The proposed stereo-video method operates without visually coded targets, decreasing the expected acquisition time to the duration of the captured videos (8 to 10 min per plot as shown in Table II). In contrast, the `monocular-video` approach [31] requires additional time to set up the visually coded targets, equal to 25 min for our study area. While the stereo-video approach significantly speeds up the acquisition, its processing requirements are higher (see Table II). Compared with the `monocular-video` approach, the proposed method takes approximately three times longer to align the video frames and two times longer to generate a point cloud. The more extended computation is due to the amount of information, which doubles when stereo cameras are involved.

IV. Discussion

A. Stereo Videogrammetry

In this study, we have presented a low-cost stereo-video system for 3-D reconstruction of forest sites. While other stereo-based data acquisition and point cloud reconstruction systems

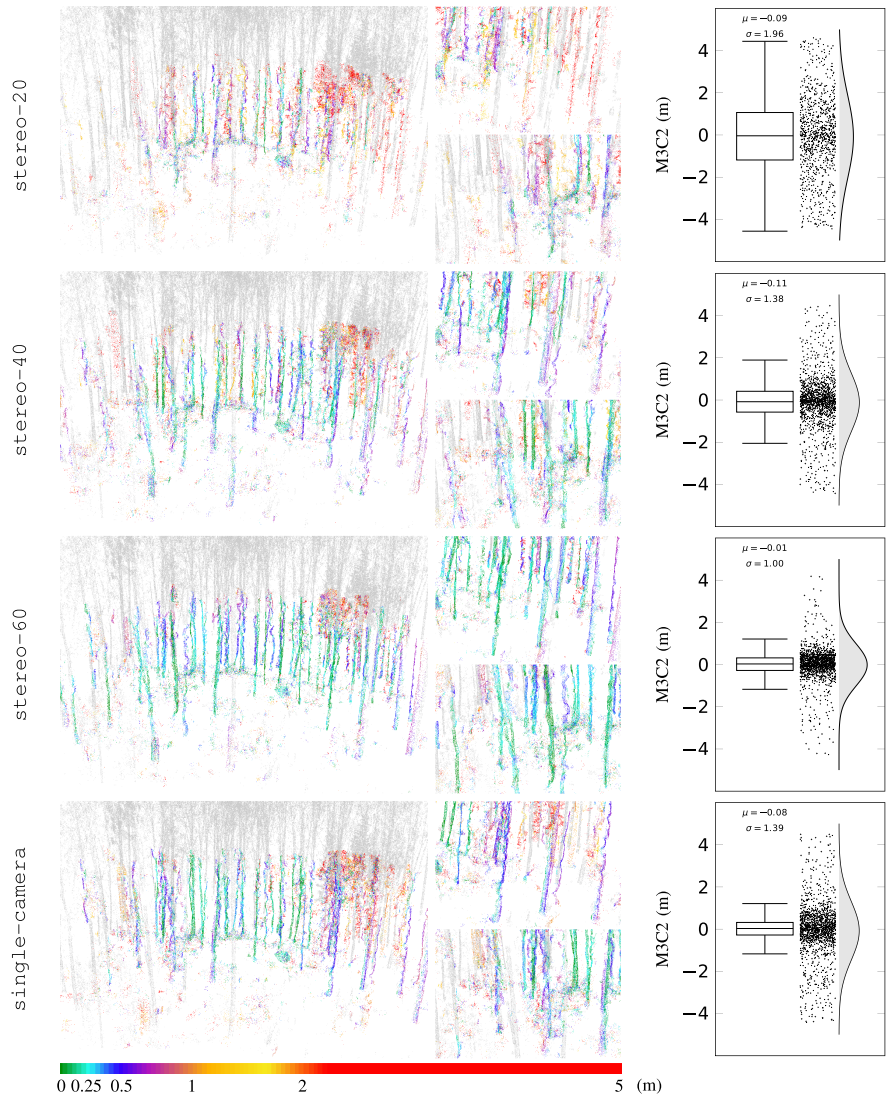

Fig. 7. M3C2 metric on point clouds containing tree stems only. The first two columns visualize the cloud-to-cloud distance change as color maps. Green indicates close-to-zero errors, whereas red indicates large errors (greater than 2 m). The parts of the TLS point cloud that were not recovered by a given videogrammetric approach are shown in gray. The third column shows the M3C2 error distributions and their mean and standard deviation values.

TABLE II
ACQUISITION AND PROCESSING TIMES FOR ALL VIDEOGRAMMETRIC APPROACHES IN HH:MM FORMAT

		stereo-20		stereo-40		stereo-60		monocular-video*	
	Acquire	Align	Generate	Align	Generate	Align	Generate	Align	Generate
Plot 1	00:08+t*	12:49	06:06	11:56	04:38	11:53	05:12	03:56	02:12
Plot 2	00:09+t*	14:58	05:28	15:09	06:57	14:57	05:23	04:28	02:32
Plot 3	00:10+t*	23:24	08:28	23:32	07:35	23:09	07:47	06:58	02:54

The additional time for distributing visually coded targets T equals 00:25 and applies only to methods marked with *.

have been introduced, e.g., [47], [48], they rely on visual simultaneous localization and mapping and aim at real-time data collection and sparse 3-D reconstruction. Our methodology is different because we prioritize quality over the real-time generation of point clouds. Instead of a mobile mapping system, we solely focus on exploiting stereo-video information.

Past research on photogrammetric stereo reconstruction [49] has considered the dual fish-eye cameras of the Ricoh Theta sensor as stereo cameras, even though they are attached to the same support. Unlike in [49], our stereo system comprises two spherical cameras instead of just one, expanding the visual information and enabling better scene coverage. Moreover, the two lightweight cameras in our acquisition equipment are ideal for use in challenging forest conditions. This contrasts other solutions for forest applications that demand heavy camera rigs with multiple high-end cameras [15], [17], which can be

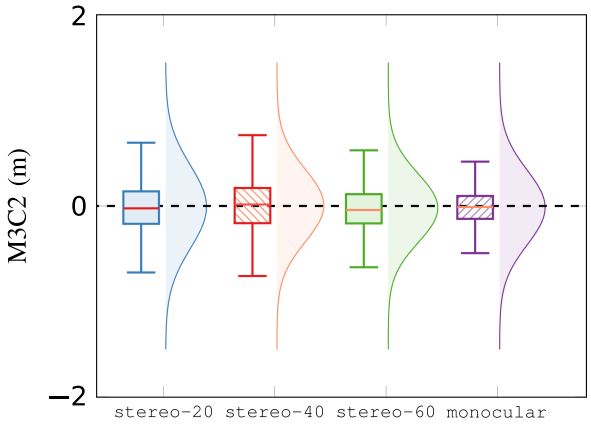

Fig. 8. M3C2 errors computed for the ground points of the videogrammetric point clouds for Plot 1. The M3C2 error distributions are shown as box plots and density curve distributions.

cumbersome and expensive. Our approach is practical, affordable, and efficient.

In our study, the stereo-video approach with a baseline of 60 cm outperforms the monocular-video approach. Yet, while using a stereo baseline of 40 cm, we achieved the same level of reconstruction precision as with the `monocular-video` approach. However, the quality of the point cloud reconstructed from a monocular video sequence depends heavily on three factors: 1) the number of control points (represented by the visually coded targets), 2) their distribution, and 3) their proper detection by the software. Ideally, more control points detected in a more distributed manner lead to better precision, similar to the requirements for aerotriangulation in classical photogrammetry [50]. The first factor is always considered a compromise between precision and time spent in the field [33], while the second factor applies to all 3-D registration processes, being of particular importance in photogrammetry due to possible model distortions [34], [51]. Meanwhile, the third factor is directly related to the image resolution [52].

Finally, the `monocular-video` approach requires half as many spherical video frames as the stereo-video approach to generate the point cloud. While this results in greater computational efficiency, less data is utilized, potentially compromising the dense matching performance and resulting in lower precision and fewer details in the point cloud (see Section III).

B. Influence of Stereo Baseline

Our results show that increasing the baseline length positively affects the quality of the point clouds generated from stereo-video sequences. We achieved the highest point cloud reconstruction quality when using a baseline of 60 cm (`stereo-60` approach). This setup outperformed the stereo-video approaches with 20 and 40 cm baselines as well as the `monocular-video` approach. In theory, using a larger baseline could result in even higher precision. Hasegawa et al. [53] studied optimal baseline values in the context of aerial stereo imagery. The authors demonstrated that the best digital elevation model can be created for a baseline-to-height ratio that does not exceed 1. However, the study has also found that large baseline values can reduce accuracy because the intersection angle formed between the two camera positions and the object is the principal factor determining quality. In a terrestrial close-range photogrammetry setting, this problem is less apparent due to the reduced camera-to-object distance. With our stereo-camera system, we target applicability to various forest conditions. While increasing the baseline further may enhance the accuracy of the 3-D reconstruction, it may not always be practical. In a heterogeneous forest environment, relying on large baselines can harm the acquisition experience by causing unwanted interaction of the equipment with the forest components (e.g., forest understorey). This interference can result in a significant camera shake, causing poor video quality and, thus, negatively affecting the point cloud reconstruction. To consider these issues, in our case study, we limited the baseline to 60 cm.

C. Influence of Spherical Camera Lenses

The advantages of using spherical cameras in our method are twofold: 1) spherical camera sensors help capture more content in less time, and 2) they contribute to the successful frame orientation during the point cloud reconstruction. Spherical lenses ensured sufficient overlap between video frames needed to successfully generate point clouds, even when moving with a moderate walking speed during data acquisition. This resulted in a short data acquisition time ranging from 8 to 10 min per 50×50 m^2 plot. In contrast, Mokroš et al. [17] required a reduction in the walking speed to ensure image overlap while capturing stereo image data with their multicamera system comprising four high-end cameras. On average, that resulted in a data acquisition time of 8 min per 25×25 m^2 plot.

Furthermore, a key step in point cloud generation is the proper frame orientation, which leads to greater reconstruction precision. The frame orientation depends on the spherical stereo camera calibration. In our case, an automatic camera calibration was performed in AgiSoft Metashape. Further exploration of the calibration aspect may benefit the overall absolute accuracy of the generated point cloud and is a potential avenue for future improvement.

D. Extracting Forest Attributes

Forest conditions influence the performance of the videogrammetric approach. Although this article's focus is not on demonstrating the generazability of the stereo-video approach, we explored its potential for extracting forest attributes in three different types of forest conditions. Plot 1 represented a relatively open forest area, unlike Plots 2 and 3, where a dense understory was present. As shown in Fig. 5(b), the smallest mean tree position error of 30 cm for our `stereo-60` approach was obtained for Plot 1. In Fig. 5(a), it is evident that not all reference trees were detected in the generated point clouds. This could result from missing information caused by occlusion during data acquisition. The selected grid pattern is also a determining factor that affects the completeness of the generated point cloud [18], [25]. Fig. 5(a) shows that insufficient information (either from occlusion or incomplete acquisition pattern, or a combination of the two) may hinder the proper reconstruction of trees near the edge of the forest plots [see Fig. 1(a)].

Furthermore, the tree detection rates for Plots 2 and 3 are significantly lower than that of Plot 1, as shown in Fig. 4(a). The lower detection rate in Plots 2 and 3 could be attributed to the more complex forest conditions in these plots. Only 33% or less of the trees with a DBH of less than 10 cm were detected in the videogrammetric point clouds, as shown in Fig. 4(b). Dense forest areas pose a challenge to the acquisition process, making it difficult to capture feasible content within such areas. This leads to a significant loss of information, resulting in low matching success and an incomplete point cloud.

Given the low-cost nature of our `stereo-60` approach, having an average tree position error of 30 to 50 cm compared to the TLS data is a clear improvement over the `monocular video` approach using the same type of spherical sensor. Various geometric factors, including the quality of point cloud coregistration, may influence this error. Indeed, the ICP method is susceptible to errors caused by high rates of noise points resulting from sensor resolution and scene heterogeneity.

E. Limitations

1) Point Cloud Orientation: The videogrammetric point clouds generated using our stereo-video approach are scaled but not absolutely oriented. The orientation of the generated point clouds is arbitrary, as per standard assumption in close-range photogrammetry. In this study, we oriented the generated videogrammetric point clouds according to the reference TLS point cloud for evaluation purposes. This allowed us to successfully match the generated point cloud to the reference TLS point cloud and extract tree positions and diameters using our algorithm that operates on a verticalized point cloud. Without reference TLS data, verticality should be established using other means. For example, this can be achieved by placing a leveled vertical reference (e.g., a surveying pole) or integrating a lightweight Global Navigation Satellite System receiver. Within the context of orientation, even low-cost and low-precision receivers may be envisaged. However, in most practical cases, specific forest applications do not require point cloud orientation. For instance, performing semantic segmentation and object detection can be accomplished even if the point cloud reconstruction is not oriented. Moreover, the orientation issue can be addressed on an algorithm level by taking into account the a priori knowledge that the point cloud is not absolutely oriented.

2) Camera Sensor Resolution: The resolution of the low-cost spherical cameras used in this study is 4 K. However, once the resolution is distributed over the 360° field of view, capturing small objects in enough detail may become insufficient. Indeed, most trees with a DBH of less than 10 cm in our study area have diameter errors greater than 20%, as shown in Fig. 6(d). The latter signifies that the DBH estimation from the videogrammetric point clouds is inefficient for small trees. In contrast, the resolution of the Ricoh Theta Z1 sensor is high enough for the proposed `stereo-60` approach to reliably reconstruct trees with a DBH greater than 30 cm [see Fig. 6(d)]. However, using the proposed stereo-video setup with higher quality sensors can positively impact the 3-D reconstruction quality (especially for smaller trees) and will be investigated in the future.

F. Application Potential

Even though point clouds generated from mono- and stereo-video sequences collected by the Ricoh Theta Z1 sensor show limited potential to detect thin trees, they provided sufficient results on tree detection and DBH estimation for medium to large trees. Thus, the proposed videogrammetric approach can be a cost-effective alternative to classical photogrammetry and ground-based laser scanning when working under specific forest conditions. The videogrammetric point clouds are suitable for forest applications that do not require very high measurement precision, such as forest structure evaluation on the lower part of the canopy, habitat and forest understory characterization, and dead wood assessment. In addition, videogrammetric point clouds can complement video and image data while using artificial intelligence methods to derive target forest information, including vegetation categories and tree species.

V. Conclusion

In this study, we have 1) introduced a novel approach for the point cloud reconstruction of forest sites from spherical stereo videos and 2) presented a new piece of equipment for data acquisition, consisting of stereo spherical cameras with a known baseline. The proposed stereo-video approach is self-contained, i.e., it builds and scales point clouds using stereo-video content without visually coded targets. This reduces the acquisition time significantly. A baseline of 60 cm was the most suitable among the three baselines for obtaining feasible 3-D reconstruction precision of a forest site. This setup reduces the average tree position error to 30 cm in easier forest situations and to 50 cm in more challenging forest conditions, while keeping the mean DBH error at less than 20%. Our `stereo-60` approach outperforms the reference `monocular-video` method. However, decreasing the baseline to 40 cm drops the point cloud accuracy to a level similar to the `monocular-video` approach, albeit with the benefit of saving time during data collection. Our low-cost spherical camera sensors can capture enough content to reliably reconstruct larger trees (with a DBH of 30 cm or more) using the `stereo-60` approach. Increasing the sensor resolution and quality could potentially improve 3-D reconstruction quality for smaller trees and is an avenue for future work.

Acknowledgment

The authors would like to thank Melissa Dawes for professional language editing of the manuscript.

References

[1] D. Kükenbrink, M. Marty, R. Bösch, and C. Ginzler, "Benchmarking laser scanning and terrestrial photogrammetry to extract forest inventory parameters in a complex temperate forest," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 113, 2022, Art. no. 102999.

[2] X. Liang et al., "Terrestrial laser scanning in forest inventories," *ISPRS J. Photogrammetry Remote Sens.*, vol. 115, pp. 63–77, 2016.

[3] M. Mokroš et al., "Evaluation of close-range photogrammetry image collection methods for estimating tree diameters," *ISPRS Int. J. Geo- Inf.*, vol. 7, no. 3, 2018, Art. no. 93.

[4] A. Bornand, N. Rehush, F. Morsdorf, E. Thürig, and M. Abegg, "Individual tree volume estimation with terrestrial laser scanning: Evaluating reconstructive and allometric approaches," *Agricultural Forest Meteorol.*, vol. 341, 2023, Art. no. 109654.

[5] M. Abegg, D. Kükenbrink, J. Zell, M. E. Schaepman, and F. Morsdorf, "Terrestrial laser scanning for forest inventories–tree diameter distribution and scanner location impact on occlusion," *Forests*, vol. 8, no. 6, 2017, Art. no. 184.

[6] J. Čerňava, M. Mokroš, J. Tuček, M. Antal, and Z. Slatkovská, "Processing chain for estimation of tree diameter from GNSS-IMU-based mobile laser scanning data," *Remote Sens.*, vol. 11, no. 6, 2019, Art. no. 615.

[7] M. Forsman, J. Holmgren, and K. Olofsson, "Tree stem diameter estimation from mobile laser scanning using line-wise intensity-based clustering," *Forests*, vol. 7, no. 9, 2016, Art. no. 206.

[8] X. Liang et al., "In-situ measurements from mobile platforms: An emerging approach to address the old challenges associated with forest inventories," *ISPRS J. Photogrammetry Remote Sens.*, vol. 143, pp. 97–107, 2018.

[9] C. Cabo, S. Del Pozo, P. Rodríguez-Gonzálvez, C. Ordóñez, and D. González-Aguilera, "Comparing terrestrial laser scanning (TLS) and wearable laser scanning (WLS) for individual tree modeling at plot level," *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 540.

[10] F. Leberl et al., "Point clouds," *Photogrammetric Eng. Remote Sens.*, vol. 76, no. 10, pp. 1123–1134, 2010.

[11] M. Goesele, B. Curless, and S. M. Seitz, "Multi-view stereo revisited," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2006, vol. 2, pp. 2402–2409.

[12] L. Barazzetti, M. Previtali, and F. Roncoroni, "3D modelling with the samsung gear 360," *Int. Arch. the Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 85–90, 2017.

[13] A. Murtiyoso and P. Grussenmeyer, "Experiments using smartphone-based videogrammetry for low-cost cultural heritage documentation," in *Proc. 28th CIPA Symp., 28 Août-1er Septembre*, Université Tsinghua, Pekin, Chine, vol. 46, 2021, pp. 487–491.

[14] A. J. Benítez et al., "Multi-camera workflow applied to a cultural heritage building: Alhambra's torre de la cautiva from the inside," *Heritage*, vol. 5, no. 1, pp. 21–41, 2021.

[15] M. Forsman, N. Börlin, and J. Holmgren, "Estimation of tree stem attributes using terrestrial photogrammetry with a camera rig," *Forests*, vol. 7, no. 3, 2016, Art. no. 61.

[16] X. Liang et al., "The use of a hand-held camera for individual tree 3D mapping in forest sample plots," *Remote Sens.*, vol. 6, no. 7, pp. 6587–6603, 2014.

[17] M. Mokroš et al., "Novel low-cost mobile mapping systems for forest inventories as terrestrial laser scanning alternatives," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 104, 2021, Art. no. 102512.

[18] L. Piermattei et al., "Terrestrial structure from motion photogrammetry for deriving forest inventory data," *Remote Sens.*, vol. 11, no. 8, 2019, Art. no. 950.

[19] R. Zhu, Z. Guo, and X. Zhang, "Forest 3D reconstruction and individual tree parameter extraction combining close-range photo enhancement and feature matching," *Remote Sens.*, vol. 13, no. 9, 2021, Art. no. 1633.

[20] H. Hristova, M. Abegg, C. Fischer, and N. Rehush, "Monocular depth estimation in forest environments," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. 1017–1023, 2022.

[21] T. Yun et al., "Status, advancements and prospects of deep learning methods applied in forest studies," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 131, 2024, Art. no. 103938.

[22] A. Gruen, "Fundamentals of videogrammetry–a review," *Hum. Movement Sci.*, vol. 16, no. 2-3, pp. 155–187, 1997.

[23] M. Pollefeys et al., "Detailed real-time urban 3D reconstruction from video," *Int. J. Comput. Vis.*, vol. 78, pp. 143–167, 2008.

[24] K. Kwiatek and R. Tokarczyk, "Photogrammetric applications of immersive video cameras," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 2, no. 5, 2014, Art. no. 211.

[25] M. Pepe, V. S. Alfio, D. Costantino, and S. Herban, "Rapid and accurate production of 3D point cloud via latest-generation sensors in the field of cultural heritage: A comparison between SLAM and spherical videogrammetry," *Heritage*, vol. 5, no. 3, pp. 1910–1928, 2022.

[26] Y. Zhou, G. Gallego, H. Rebecq, L. Kneip, H. Li, and D. Scaramuzza, "Semi-dense 3D reconstruction with a stereo event camera," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 235–251.

[27] F. Remondino, "Heritage recording and 3D modeling with photogrammetry and 3D scanning," *Remote Sens.*, vol. 3, no. 6, pp. 1104–1138, 2011.

[28] L. Perfetti et al., "Fisheye photogrammetry to survey narrow spaces in architecture and a hypogea environment," in *Latest Developments Reality-Based 3D Surveying Modelling*. MDPI Books, 2018, pp. 3–28.

[29] B. Alsadik, M. Gerke, and G. Vosselman, "Efficient use of video for 3D modelling of cultural heritage objects," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. II-3/W4, pp. 1–8, 2015.

[30] Z. Sun and Y. Zhang, "Accuracy evaluation of videogrammetry using a low-cost spherical camera for narrow architectural heritage: An observational study with variable baselines and blur filters," *Sensors*, vol. 19, no. 3, 2019, Art. no. 496.

[31] A. Murtiyoso, H. Hristova, N. Rehush, and V. Griess, "Low-cost mapping of forest under-storey vegetation using spherical photogrammetry," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 48, pp. 185–190, 2022.

[32] C. Chioni, A. Maragno, A. Pianegonda, M. Ciolli, S. Favargiotti, and G. A. Massari, "Low-cost 3D virtual and dynamic reconstruction approach for urban forests: The mesiano university park," *Sustainability*, vol. 15, no. 19, 2023, Art. no. 14072.

[33] A. Murtiyoso, P. Grussenmeyer, and D. Suwardhi, "Technical considerations in low-cost heritage documentation," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci. - ISPRS Arch.*, vol. XLII-2/W17, no. 2/W17, pp. 225–232, 2019.

[34] E. Nocerino, F. Menna, F. Remondino, and R. Saleri, "Accuracy and block deformation analysis in automatic UAV and terrestrial photogrammetry - lesson learnt," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. II-5/W1, pp. 2–6, 2013.

[35] L. Barazzetti, "Network design in close-range photogrammetry with short baseline images," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. IV-2/W2, pp. 17–23, 2017.

[36] F. Dai, Y. Feng, and R. Hough, "Photogrammetric error sources and impacts on modeling and surveying in construction engineering applications," *Visual. Eng.*, vol. 2, no. 1, pp. 1–14, 2014.

[37] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *Int. J. Comput. Vis.*, vol. 27, pp. 161–195, 1998.

[38] Davinci resolve (version 18) (software), 2024. [Online]. Available: https://www.blackmagicdesign.com/products/davinciresolve,author=BlackMagicDesign

[39] P. Wilkes et al., "Data acquisition considerations for terrestrial laser scanning of forest plots," *Remote Sens. Environ.*, vol. 196, pp. 140–153, 2017.

[40] AgiSoft, "Agisoft metashape professional (version 1.4.5) (software), 2018. [Online]. Available: http://www.agisoft.com

[41] A. Murtiyoso and P. Grussenmeyer, "Documentation of heritage buildings using close-range UAV images: Dense matching issues, comparison and case studies," *Photogrammetric Rec.*, vol. 32, no. 159, pp. 206–229, 2017.

[42] CloudCompare, "Cloudcompare: 3D point cloud and mesh processing software," version 2.12, GPL software, 2023. [Online]. Available: http://www.cloudcompare.org/

[43] W. Zhang et al., "An easy-to-use airborne LiDAR data filtering method based on cloth simulation," *Remote Sens.*, vol. 8, no. 6, 2016, Art. no. 501.

[44] A.-V. Vo, L. Truong-Hong, D. F. Laefer, and M. Bertolotto, "Octree-based region growing for point cloud segmentation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 104, pp. 88–100, 2015.

[45] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[46] D. Lague, N. Brodu, and J. Leroux, "Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the rangitikei canyon (NZ)," *ISPRS J. Photogrammetry Remote Sens.*, vol. 82, pp. 10–26, 2013.

[47] A. Torresani, F. Menna, R. Battisti, and F. Remondino, "A v-slam guided and portable system for photogrammetric applications," *Remote Sens.*, vol. 13, no. 12, 2021, Art. no. 2351.

[48] D. Holdener, S. Nebiker, and S. Blaser, "Design and implementation of a novel portable 360 stereo camera system with low-cost action cameras," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 105–110, 2017.

[49] J. Caracotte, F. Morbidi, and E. M. Mouaddib, "Photometric stereo with twin-fisheye cameras," in *2020 25th IEEE Int. Conf. Pattern Recognit.*, 2021, pp. 5270–5277.

[50] Y. Lumban-Gaol, A. Murtiyoso, and B. Nugroho, "Investigations on the bundle adjustment results from SFM-based software for mapping purposes," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci. - ISPRS Arch.*, vol. XLII-2, pp. 623–628, 2018.

[51] E. Lachat, T. Landes, and P. Grussenmeyer, "Comparison of point cloud registration algorithms for better result assessment–towards an open-source solution," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLII-2, pp. 551–558, 2018.
[52] P. Sapirstein, "Accurate measurement with photogrammetry at large sites," *J. Archaeological Sci.*, vol. 66, pp. 137–145, 2016.
[53] H. Hasegawa, K. Matsuo, M. Koarai, N. Watanabe, H. Masaharu, and Y. Fukushima, "Dem accuracy and the base to height (B/H) ratio of stereo images," *Int. Arch. Photogrammetry Remote Sens.*, vol. 33, no. B4/1; PART 4, pp. 356–359, 2000.

Hristina Hristova received the B.Sc. degree in applied mathematics from the University of Sofia, Sofia, Bulgaria, in 2013, and the M.Sc. degree in distributed computing and computed vision, in 2014, and the Ph.D. degree in the field of computer vision and computer graphics in 2017 from the University of Rennes 1, Rennes, France, respectively.

She did a Postdoc in virtual reality and 5G networks with IMT Atlantique, Rennes, and in virtual reality and artificial intelligence (AI) solutions for forest applications, Swiss Federal Institute for Forest, Snow, and Landscape Research (WSL), Birmensdorf, Switzerland, where she is currently working as a scientific collaborator. Her research interests include virtual reality, AI and machine learning for forestry, remote sensing, and photogrammetry.

Arnadi Murtiyoso received the B.Sc. degree in geodesy and geomatics from the Bandung Institute of Technology, Bandung, Indonesia, in 2011, the French engineering diploma in topography and surveying from INSA Strasbourg, Strasbourg, France, in 2016, and the Ph.D. degree in photogrammetry and geomatics from the University of Strasbourg, Strasbourg.

He is currently working as a Postdoc with the Forest Resources Management Research Group, ETH Zurich, Zurich, Switzerland. His research interests include close-range 3D reconstruction techniques and data visualization for forest environments.

Daniel Kükenbrink received the M.Sc. and Ph.D. degrees in geography from the University of Zürich, Zürich, Switzerland, in 2014 and 2019, respectively.

Since 2019, he has been working with the Swiss National Forest Inventory (NFI), Swiss Federal Institute for Forest, Snow and Landscape Research, Birmensdorf, Switzerland. His research interests include the assessment of 3-D forest structures using close-range remote sensing techniques, interaction of forest structure and light distribution within forest canopies, and evaluates the operational inclusion of close-range remote sensing with the framework of the Swiss NFI.

Mauro Marty received the M.Sc. degree in geography from the University of Bern, Bern, Switzerland in 2012.

Since 2012, he has been working with the Swiss Federal Institute for Forest, Snow, and Landscape Research, Birmensdorf, Switzerland, where he is currently a Technical Staff Member.

Meinrad Abegg received the M.Sc. degree in forest sciences and the Ph.D. degree in geography from ETH Zürich, Zürich, Switzerland, in 2003 and 2020, respectively.

Since 2004, he has been working with the Swiss National Forest Inventory (NFI), Swiss Federal Institute for Forest, Snow, and Landscape Research (WSL), Birmensdorf, Switzerland, working on different subjects along the workflow from field measurements to outreach products with NFI results. His research interests include the application of terrestrial laser scanning for forest inventories, inventory statistics, NFI data processing, and the development of field methods for forest inventories.

Christoph Fischer received the M.Sc. degree in tropical and international forestry and the Ph.D. degree in forestry from Georg-August-Universität Göttingen, Göttingen, Germany, in 2008 and in 2011, respectively.

Since 2011, he has been working with the Swiss National Forest Inventory (NFI), Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland. Since 2020, he has been the Head of the Research Group Scientific Service NFI. His research interest include methods for field surveys, remote sensing applications, inventory statistics, and the provision of forest services, especially recreation.

Verena C. Griess received the diploma course in forestry engineering and the master's degree in forest and wood sciences in 2008, and the Ph.D. degree from the Technical University of Munich, Munich, Germany, in 2012.

From 2014 to 2021, she was an Assistant Professor with the University of British Columbia, Vancouver, BC, Canada. Since January 2021, she has been a Full Professor of Forest Resources Management, ETH Zurich, Zurich, Switzerland. Her research interests include multiple disciplines encompassing various aspects of forestry, modeling, economics, decision theory, and risk management.

Nataliia Rehush received the M.Sc. degree in forestry and the Ph.D. degree in forest sciences from Ukrainian National Forestry University, Lviv, Ukraine, in 2010 and 2015, respectively.

Since 2016, she has been working with the Swiss National Forest Inventory (NFI), Swiss Federal Institute for Forest, Snow, and Landscape Research (WSL), Birmensdorf, Switzerland. Her research interests include remote sensing for forest applications and machine learning.