A Lightweight Leddar Optical Fusion Scanning System (FSS) for Canopy Foliage Monitoring

Zhouxin Xi[iD]**, Christopher Hopkinson *, Stewart B. Rood, Celeste Barnes, Fang Xu, David Pearce and Emily Jones**

Department of Geography, University of Lethbridge, Lethbridge, AB T1K 3M4, Canada
* Correspondence: c.hopkinson@uleth.ca; Tel.: +1-403-332-4586

check for updates

Abstract: A growing need for sampling environmental spaces in high detail is driving the rapid development of non-destructive three-dimensional (3D) sensing technologies. LiDAR sensors, capable of precise 3D measurement at various scales from indoor to landscape, still lack affordable and portable products for broad-scale and multi-temporal monitoring. This study aims to configure a compact and low-cost 3D fusion scanning system (FSS) with a multi-segment Leddar (light emitting diode detection and ranging, LeddarTech), a monocular camera, and rotational robotics to recover hemispherical, colored point clouds. This includes an entire framework of calibration and fusion algorithms utilizing Leddar depth measurements and image parallax information. The FSS was applied to scan a cottonwood (*Populus* spp.) stand repeatedly during autumnal leaf drop. Results show that the calibration error based on bundle adjustment is between 1 and 3 pixels. The FSS scans exhibit a similar canopy volume profile to the benchmarking terrestrial laser scans, with an r^2 between 0.5 and 0.7 in varying stages of leaf cover. The 3D point distribution information from FSS also provides a valuable correction factor for the leaf area index (LAI) estimation. The consistency of corrected LAI measurement demonstrates the practical value of deploying FSS for canopy foliage monitoring.

Keywords: LiDAR; Leddar; sensor fusion; structure from motion; monocular camera; terrestrial laser scanning; leaf area index; canopy monitoring

1. Introduction

Monitoring in-situ canopy variables, such as leaf area index (LAI) from the ground is valuable for developing canopy light interception and biomass growth models for forests [1,2]. Passive optical sensors, such as digital cameras, are cost-effective ground-based tools, and have been applied to monitor canopy variables, such as LAI, and the fraction of absorbed, photosynthetically active radiation (fPAR) [3]. However, the passive optical approach provides limited accuracy due to the 3D heterogeneity of foliage distribution. For example, the LAI estimated from the Beer–Lambert geometric-optical model [4,5], also termed effective LAI, is usually 55%–65% of the true LAI [6]. The constraint of sensing tools inevitably leads to sophisticated tuning efforts for the geometric-optical model, such as the introduction of gap size distribution, clumping factor, and needle-to-shoot area ratio [7]. The emerging terrestrial laser scanning (TLS) technology significantly mitigates the LAI characterization problems. The 3D datasets from TLS not only enable straightforward estimation of canopy model variables, including leaf angle distribution (LAD) [8,9], clumping index [10], canopy foliage profile [11], gap fraction [12], and plant area volume density (PAVD) [13,14], but have also led to the development of more accurate canopy models. For example, the non-randomness of leaf distribution, conventionally described using the clumping index or gap size distribution, can be explicitly modeled in path length distribution equations [15,16]. Basically, the path length distribution (PATH) model relates the light extinction degree with the detailed optical paths extractable from 3D point clouds. It is adapted from

the foliage profile equation in [17]. The LAI variable can be estimated from the PATH model with high accuracy and stability benchmarked by the true LAI [7,15]. Therefore, capturing 3D information has substantial potential for modeling canopy variables.

Conventional TLS sensors have shown impressive measurement precision and reliability of capturing 3D canopy information, yet lack portability and affordability for widespread use at the landscape level. An alternative solution is to integrate low-cost LiDAR sensors into an existing ground sensor network for broad-scale monitoring purposes. Manufacturers such as Faro and Leica produce portable terrestrial or mobile LiDAR at a price range of 4000 to 20,000 USD with moderate frequency (20–300 s^{-1}) and centimeter-level resolution. Those sensors are not sufficiently cost effective, power saving, compact, or flexible for widespread 3D biomass monitoring. In recent years, tiny LiDAR scanners, such as those available from Velodyne, Ouster, Hokuyo, SICK, Ibeo, and Scanse have entered the market with a price level between 100 and 4000 USD. Most tiny scanners have a limited field of view (FOV) and low point detection frequency. Many low-cost tiny scanners scan in only 2D, recording laser distance returns with a spinning mirror or motor. Thus, multiple scanlines are required to produce sufficient vertical points, whereas low frequencies can cause serious 3D distortions from fast-moving platforms or targets [18].

A multi-segment (sometimes referred to as multi-beam) laser scanner may provide a balanced choice of budget and frequency. Instead of repetitive scanning in the vertical direction, a multi-segment scanner relies on detection arrays to record multiple distances instantly. Each segment from a detection array usually has high detection frequency over 50 s^{-1}. Among the multi-segment scanners, an LED-based LiDAR from LeddarTech, known as Leddar (light emitting diode detection and ranging), stands out as an economical choice, with 16 segments, 100 s^{-1} frequency, and a thousand-dollar cost. A Leddar sensor implements patented algorithms to estimate traveling distances of each pulse emitted from an LED light source and detected by an array of 16 PIN photodiodes [19]. Each segment corresponds to a solid angle of approximately 2.8° × 7.5° and the field of view for 16 segments is customizable between 9° and 95°. The sensor can record multiple distance returns from multiple objects at different distances. Leddar's capability of rapid data acquisition and multiple object detection has enabled growing applications in canopy detection [20], autonomous driving [21,22], traffic analysis [23,24], parking assistance [25], and drone altitude estimation [26,27]. However, due to limited FOV and sparse segments, the Leddar sensor alone cannot compete with conventional static TLS for detailed 3D canopy modeling or geometric mensuration.

Adding a twin-axis rotational robot to the Leddar sensor can be a cost-effective solution to expand its FOV and enhance point clouds. The boresight of the integrated system, however, needs to be calibrated in order to deliver precise and consistent 3D data or point clouds. One type of LiDAR calibration method is statistical correction based on a known target. For example, Bohren et al. [28] correct 3D ground point clouds from SICK and Velodyne scanners based on the planarity constraint of the ground. A more rigorous calibration method is to model the physical relationship between the LiDAR system and calibration targets. For example, Muhammad and Lacroix [29] use a planar target to calibrate five intrinsic parameters of a Velodyne HDL-64E S2 system mounted on a static rotator, including two segment angles and three origin parameters. Atanacio-Jiménez et al. [30] expand the calibration target to be a room of five planes, and calibrate intrinsic and extrinsic parameters of a Velodye HDL-64E. Zhu and Liu [18] estimate the origin and orientation of the HDL64E S2 sensor by aligning point clouds based on pole-shaped features. Several other studies provide convenient physical calibration approaches without a requirement for measuring reference target coordinates. For example, Levinson and Thrun [31] propose a global calibration method for 192 orientation or distance parameters generated from a moving trajectory of a Velodyne HD-64E S2 sensor. Point clouds are self-calibrated by maximizing local planarity without a need for a deliberate reference target. Similarly, Sheehan et al. [32] proposed an entropy-based self-calibration method, maximizing the sharpness of point clouds for three SICK LMS-151 laser scanning units.

The above calibration methods require accurate position and orientation data from external sources, such as a wheel encoder, GPS, or the inertial measurement unit (IMU), and also dense points to support point cloud alignment and registration. These requirements are usually not easily met in a low-cost scanning system with Leddar and a rotational motor. An alternative solution is to integrate a camera sensor. The extrinsic parameters, such as pose and origin, can be estimated using photogrammetric methods. The predominant approach from existing studies is to re-project LiDAR point clouds to a 2D plane and co-register with an image based on corresponding features [33–38]. The calibration accuracy from these studies substantially depends on highly dense point clouds, which a low-cost Leddar sensor is unable to produce. Among the few studies on calibrating sparse LiDAR data, Debattisti et al. [39] places focus on edge points of artificial targets visible from both an image and a SICK LMS221 sensor. Debattisti, Mazzei, and Panciroli [39] also point out that point clouds from low-cost LiDAR usually lead to laborious scanning in pursuit of sufficient corresponding points from both the image and point cloud. Considering the characteristics of a low-cost scanning system, this study will utilize a planar calibration target to avoid the prerequisite for dense point clouds and also integrate a camera to provide extrinsic pose estimation.

Adding a camera sensor to the scanning system not only satisfies the calibration and alignment need but also provides useful texture details. A question of interest is how to integrate the texture information from a camera with the sparse point clouds from a Leddar scanning system, in order to produce densely colored point clouds. Indeed, a variety of existing literature, including Hartley and Zisserman [40] in particular, already illustrates the feasibility of reconstructing dense 3D point clouds from stereo or multiple images directly, without LiDAR distance. However, in regard to our small scanning system with one rotational camera, the short movement baseline would lead to poor 3D reconstruction quality. Assuming the target is far from the camera, the error of depth estimate (ΔD) is related to the error of stereo matching between two images (Δx) in Equation (1):

$$\Delta x \;=\; \frac{fd\Delta D}{\mu D^2} \tag{1}$$

Assuming the baseline *d* of the small scanning system is about 3 cm, the focal length *f* is 3.6 mm, pixel size *μ* is 2.8 μm, and the target distance *D* is 20 m, then a small Δx of 0.1 pixel will lead to a ΔD of 1 m. We can conclude that using the monocular camera alone without Leddar point clouds in this case cannot produce an accurate point cloud. Although it is feasible to integrate both Leddar depth and camera parallax in a bundle adjustment equation to reconstruct 3D points, in our ill-posed case, the camera has a highly limited contribution to recovering depth (Z) information. The main role of the camera in the bundle adjustment is to regularize the point clouds in the 2D plane (XY). After fine bundle adjustment, the camera can also be useful for filling gaps between the segments from the interpolation point of view. The interpolation methods are many (e.g., De Silva et al. [41] match the resolution between image and point clouds using a Gaussian process model for dense 3D recovery).

The objective of this study is to (1) configure a compact and low-cost fusion scanning system (FSS) including a multi-segment Leddar, monocular camera, and rotational robotics; and (2) propose an entire framework of calibration and fusion algorithms that produce dense colored point clouds covering a hemispherical view for 3D canopy monitoring. The specific technical refinements are (1) the addition of a kinematic motion constraint to the spatiotemporal bundle adjustment equations for the FSS calibration, and (2) the iterative optimization of monocular camera parallax and Leddar depth under ill-posed conditions. These technical contributions enable a cost-effective FSS, providing dense point clouds with rapid canopy information, such as gap fraction and LAI, useful for monitoring foliage biomass status and changes. Most parts of the framework are automatic and aimed to reduce potential manual intervention. Undoubtedly, the development of diverse lightweight sensors, particularly LiDAR, has been expanding the manner in which our environment can be sampled and digitized. It is time to tailor the tool selection to the application demand, instead of simply pursuing ultimate resolution. For example, as Table 1 indicates, in the canopy monitoring situation, with emphasis on

sensor scalability and durability, the FSS can stand out as a flexible and balanced option among the four canopy sampling tools. We expect our work could promote more research attention to the lightweight scanning systems as a cost-effective option for 3D environmental mapping and monitoring.

Table 1. Qualitative ranking of advantages among terrestrial laser scanning (TLS), digital hemispherical photography (DHP), fusion scanning system (FSS), and sweeping 2D LiDAR for canopy sampling. Ranks are denoted as ++, + and - in descending order.

|	TLS	DHP	FSS	2D LiDAR
Spatial resolution	++	++	+	-
Detection range	++	+	+	-
Equipment affordability	-	-	+	++
Operative efficiency	+	++	+	+
3D measurement accuracy	++	-	+	+
Portability and scalability	-	-	+	++
Repeatability and durability	-	-	+	+

2. Materials and Methods

2.1. Hardware Customization and Data Processing Framework

The Leddar optical FSS consists of two sensors: a 16-segment Leddar sensor (Leddar M16) and a web camera sensor in a 3D printed enclosure ($13.3 \times 9.1 \times 4.1$ cm). The two lightweight sensors (<300 g) sit on top of a tilting arm (DDT560 Direct Drive Tilt) and a panning base, as shown in Figure 1. Beside and beneath the pan-tilt arms are rotational servos, which drive the pan-tilt movement and determine the angular resolution and span of the rotation. Specifically, the tilting servo (Hitec HS-5485HB) is a standard digital servo that rotates between 0 and 118° with a highest resolution of 0.6°. The pan servo (Dynamixel MX-12W) can have 360° rotation with 0.08° resolution and can feedback rotation angles in real time. Specifications of Leddar, camera, and servos are provided in Table 2. Camera video and Leddar distance data are collected and stored by a Raspberry Pi, and servo rotations are manipulated by an Arduino Mega 2560 board. The detailed connections between sensors, pan-tilt robotics, and electronic controllers are shown in Figure 2.
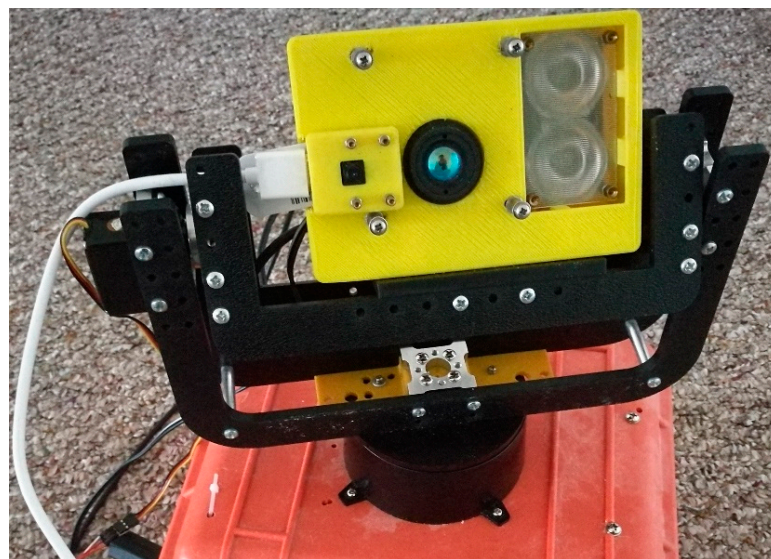
Figure 1. A fusion scanning system (FSS) with light emitting diode detection and ranging (Leddar) and monocular camera sensors.
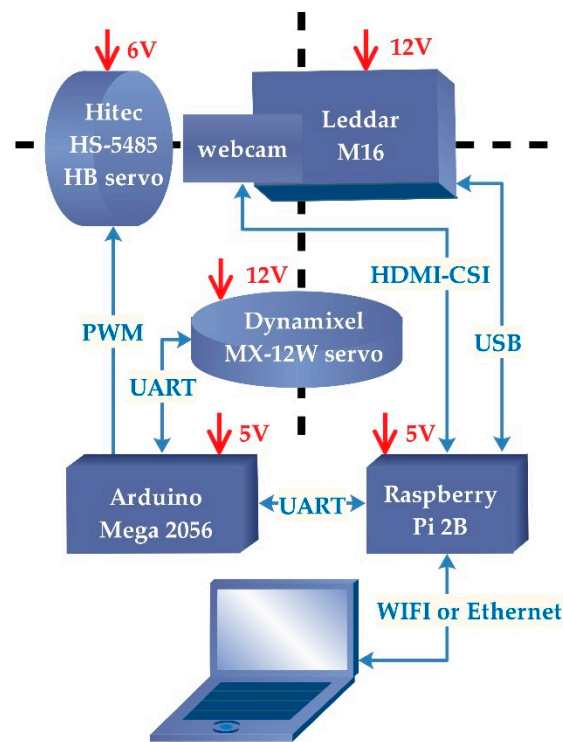
Figure 2. Hardware components and connections of the FSS. PWM: Pulse width modulation. UART: Universal asynchronous receiver/transmitter. CSI: Camera serial interface.

Table 2. Major hardware specifications.

Camera	Leddar	Tilt Servo	Pan Servo
OmniVision OV5647	M16 module	Hitec HS-5485HB	Dynamixel MX-12W
FOV: 54° × 41°	Distance: 0 to 50 m	Max angle: 118°	Max angle: 360°
Lens: f = 3.6 mm, f/2.9	Frequency: ≤100 s^{-1}	PWM: 750–2250 µs	Steps: 4096
Calibration: no IR	Wavelength: 940 nm	Deadband: 8 µs	Resolution: 0.088°
Application: IR filter	Power: 12/24 V, 4 W	Power: 4.8–6.0 V	Voltage: 12 V

To scan a wide field of view (e.g., hemispherical view), the pan-tilt system rotation follows a common raster scanning scheme: setting a fixed tilt angle for one horizontal scanline rotation and changing the tilt angle for the next. For each horizontal scanline, recordings from the Leddar, the camera, and the servo are stored asynchronously in separate files by the Raspberry Pi. The Leddar sensor outputs timestamp, segment distance, echo amplitude, and the echo quality index with an updating frequency of 100 s^{-1}. The Raspberry Pi camera captures 720p video with a rate of 25 frames per second (FPS). The rotation angle from the pan servo is saved every 4°. The Leddar and pan angle readings are synchronized based on their millisecond-level timestamps tagged by the Raspberry Pi. The camera videos do not have timestamps, and their timing is inferred from motion detection.

Constructing the hardware of the FSS is technically straightforward. The system components are inexpensive and uncomplicated compared to most commercial LiDAR scanning systems, yet the fusion of different low-resolution data sources to generate dense point clouds is the primary challenge. The framework of our multi-source data fusion is illustrated in Figure 3, including: (1) mapping discrete Leddar distances onto individual video frames to create "3D pixels," (2) aligning video frames globally to cover a panoramic field of view, and (3) adjusting frame alignment and extrapolating 3D point clouds based on the "3D pixels."
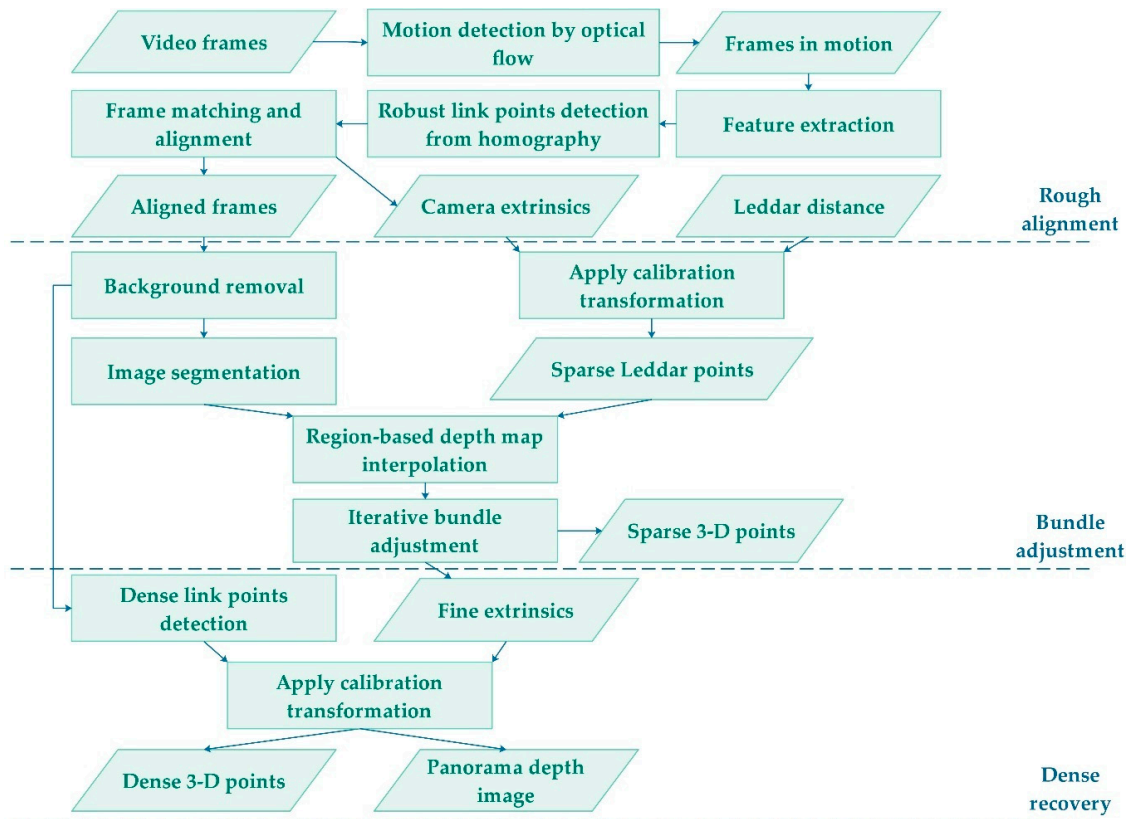
Figure 3. Framework of point cloud recovery from monocular camera and sparse Leddar segments.

2.2. Coordinate System Conversion for Calibration

Image frames decomposed from camera video do not have timestamps directly linkable to Leddar timestamps. It is necessary to match the camera motion with the Leddar motion and assign timestamps to camera frames. The start and stop of camera motion are detected by optical flows from neighboring frames: frames with average pixel velocity above a threshold of 0.6 pixels are considered moving. With start and stop timestamps, all the timestamps of the moving frames can be recovered based on linear interpolation.

The synchronization of camera and Leddar enables one-to-one mapping between Leddar distances and camera frames. We need to further locate Leddar footprints on the pixels of each video frame. This is solved through calibration of the FSS, or specifically, determining the unknown boresight parameters that convert the Leddar coordinate system to the camera coordinate system. Calibrating the FSS is only needed once, which enables the step of applying calibration transformation in the framework in Figure 3.

Our reference coordinate system (RCS) is a right-handed Cartesian coordinate system. It has the same units as the world coordinate system (WCS), with its origin at the camera optical center, the x axis along the transverse direction of the image plane, the y axis along the longitudinal direction of the image plane, and the z axis along the camera's optical view direction. The target 3D coordinates $\overline{P}_t$ in RCS can be parametrized in Equation (2):

$$\overline{P}_t = \begin{pmatrix} T \\ 0 \\ 0 \end{pmatrix} + (D_{it} + D_b) \begin{pmatrix} \cos(\theta_i)\sin(\psi_i) \\ \sin(\theta_i)\sin(\psi_i) \\ \cos(\psi_i) \end{pmatrix} \tag{2}$$

where t is a specific time point, T the horizontal location of Leddar optical center in RCS, polar angles θ_i and ψ_i the orientation of ith Leddar segment in RCS, D_{it} the distance measurement of ith segment,

D_b the bias of distance measurement, and $\overline{P}_t$ the target 3D coordinates in RCS. We consider T, D_b, θ_i, and ψ_i to be constant during pan-tilt rotation to represent no relative movement between the Leddar and camera. Assuming all the segments are equiangular, angle ψ_i can be presented by an arithmetic sequence parametrized with ψ_0 and ψ_Δ (Equation (2)).

The real-world coordinates of target P can be converted from $\overline{P}_t$ in RCS through the camera extrinsic parameters in Equation (3):

$$P = R_t \overline{P}_t + T_t \tag{3}$$

where R_t is the rotation matrix, and T_t the translation vector. R_t can be characterized by Euler angles (α_t, β_t, and γ_t) following $z - y - x$ rotation order. Assuming the initial rotation matrix is R_0 or Euler angles (α_0, β_0, and γ_0), the temporal change of R_t during a horizontal scan can be represented by rotation matrix R_w in Equation (4):

$$R_t = R_0 R_w \tag{4}$$

$$R_w = \begin{pmatrix} 1 - 2(u_y u_y + u_z u_z) & 2(u_x u_y - u_w u_z) & 2(u_w u_y + u_x u_z) \\ 2(u_x u_y + u_w u_z) & 1 - 2(u_x u_x + u_z u_z) & 2(u_y u_z - u_w u_x) \\ 2(u_x u_z - u_w u_y) & 2(u_y u_z - u_w u_x) & 1 - 2(u_x u_x + u_y u_y) \end{pmatrix};$$

$$T_t = \begin{pmatrix} X_{ct} \\ Y_{ct} \\ Z_{ct} \end{pmatrix} = \begin{pmatrix} X_{c0} + a_1\phi_x + a_3\phi_z - (a_{01}\phi_x + a_{03}\phi_z) \\ Y_{c0} + b_1\phi_x + b_3\phi_z - (b_{01}\phi_x + b_{03}\phi_z) \\ Z_{c0} + c_1\phi_x + c_3\phi_z - (c_{01}\phi_x + c_{03}\phi_z) \end{pmatrix};$$

where

$$\begin{pmatrix} u_x \\ u_y \\ u_z \\ u_w \end{pmatrix} = \begin{pmatrix} \sin(\phi_\alpha)\cos(\phi_\beta)\sin(\omega_t/2) \\ \cos(\phi_\alpha)\sin(\omega_t/2) \\ \sin(\phi_\alpha)\sin(\phi_\beta)\sin(\omega_t/2) \end{pmatrix};$$

$$R_t = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}, R_0 = \begin{pmatrix} a_{01} & a_{02} & a_{03} \\ b_{01} & b_{02} & b_{03} \\ c_{01} & c_{02} & c_{03} \end{pmatrix}.$$

where ϕ_α and ϕ_β define the horizontal rotation axis and w_t is the horizontal rotation angle measured by the servo. The camera optical center T_t also slightly moves during the horizontal scan, whose temporal change can be parameterized by its initial position (X_{c0}, Y_{c0}, and Z_{c0}) and rotation origin (ϕ_x and ϕ_z) in Equation (4).

Solving the calibration Equations (2)–(4) requires measurement of P and D_{it} from the same target point. However, since the exact Leddar point is invisible from the web camera, it is impossible to measure the exact 3D coordinates for P in the real world. Instead, we can reduce the requirement of the 3D P by finding a planar target with constant Z values (Z_0), and arbitrary X and Y values. Therefore, combining Equations (3) and (4), $\overline{P}_t$ and in RCS should satisfy a planar constraint in Equation (5):

$$\begin{pmatrix} c_1 & c_2 & c_3 \end{pmatrix} \overline{P}_t + Z_{ct} = Z_0 \tag{5}$$

where Z_{ct} is the Z component of T_t. With Equations (2) and (5), the only two required measurements are D_{it} at multiple time points and Z_0 of the planar target.

The above four equations form a set of nonlinear calibration equations with nine unknown intrinsic terms (T, D_b, θ_0, ψ_0, ψ_Δ, ϕ_α, ϕ_β, ϕ_x, and ϕ_z) and six initial extrinsic terms (X_{c0}, Y_{c0}, Z_{c0}, α_0, β_0, and γ_0). Our solution is iterative. The initial extrinsic terms X_{c0}, Y_{c0}, Z_{c0}, α_0, β_0, and γ_0 are solved using least-square regression of camera collinearity equations (Equation (6)):

$$x_t = K\overline{P}'_t = KR_t^T(P' - T_t) \tag{6}$$

$$K = \begin{pmatrix} \frac{f}{\mu} & 0 & x_0 \\ 0 & \frac{f}{\mu} & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

given additional measurements of pixel coordinates $x_{t=0}$ and the corresponding world coordinates P' (details in Section 2.3). The R_t^T in Equation (6) denotes transposed R_t in Equation (3). For simplicity, the camera intrinsic parameters in Equation (6) are fixed, including camera focal length f, pixel size μ, half image width x_0, and half height y_0 in pixels. Lens distortion is not considered in this study. Combining Equations (4) and (6), both extrinsic terms (X_{c0}, Y_{c0}, Z_{c0}, α_0, β_0, and γ_0) and intrinsic terms (ϕ_α, ϕ_β, ϕ_x, and ϕ_z) can be inferred by non-linear least-square regression, and further substituted into Equations (2), (3), and (5) to finally estimate the Leddar's intrinsic parameters (T, D_b, θ_0, ψ_0, and ψ_Δ).

After knowing all intrinsic and extrinsic parameters, locating Leddar points on camera images is feasible using Equations (2) and (6). It is also possible to roughly estimate a Leddar point P given D_{it} and w_t from pan-tilt angles based on Equations (2)–(4). Instead of using pan-tilt angles, we can also rely on camera global alignment in the following section for more precise inference of R_t and T_t, thus estimate P from Equation (2).
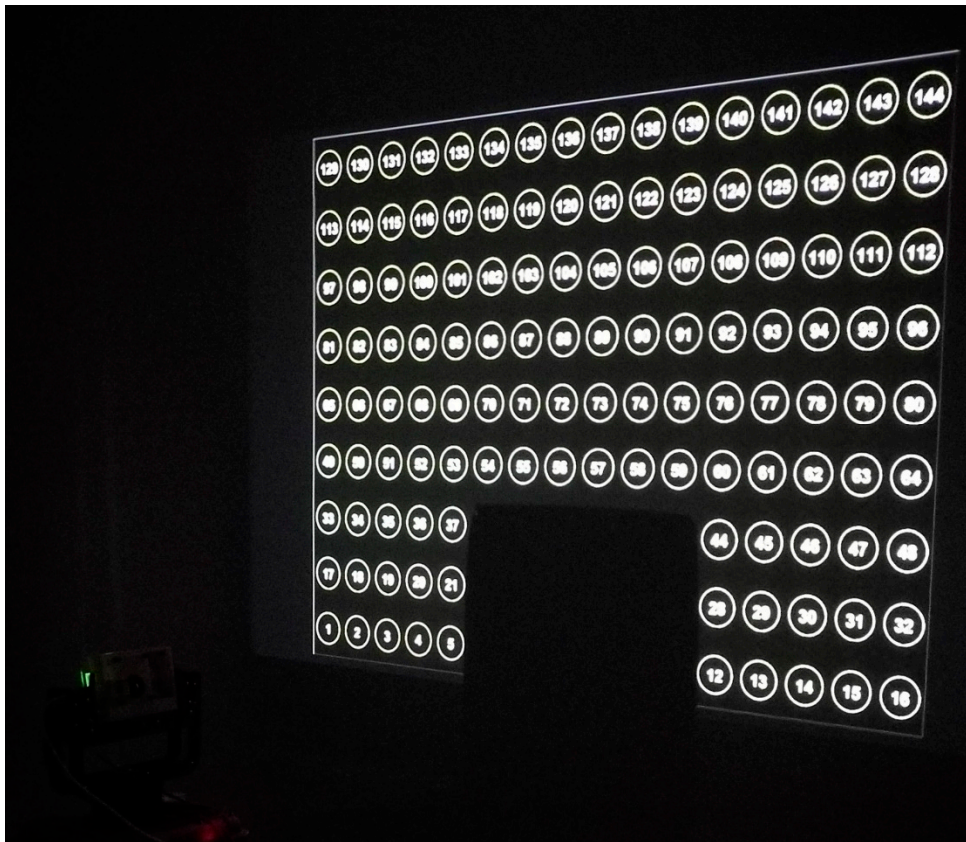
2.3. Calibration Experiment

This section presents an example of system calibration, with a flat wall ($Z_0 = 0$) being our calibration target (Figure 4). The left bottom corner of the wall was defined as the WCS origin. The sensors scanned the wall with a fixed tilt angle of about 15° and continuous horizontal angle from 50° to 140° (2.4° per second). A total of 70 frames were subsampled from the video with equal intervals for calibration. On the front wall was a 16 × 9 grid from an optical projector, for the purpose of calibrating the camera extrinsic parameters in Equation (6). As mentioned, the calibration equation (Equation (6)) required measuring the WCS coordinates P' of each projected circle center and extracting the corresponding pixel coordinates x_t from camera frames. We used a grid of circular targets instead of a chessboard pattern to be more robust with edge detection error during the automatic extraction of x_t from camera images. Since the camera had a limited field of view of around 50°, not all circles appeared in the camera images. Therefore, an ID number was assigned to each circle to help link x_t and P' automatically.

Assuming the projector lens had no distortion, we manually measured the $P's$ of the four corner circle centers on the wall and then applied bilinear interpolation to get $P's$ of all the 144 circle centers. The process of extracting x_t was challenging because circle projection on the wall from the camera's viewing perspective exhibited elliptical shapes. Extracting elliptical centers is more difficult than extracting circular centers. We adopted the characteristic number ellipse detector (CNED) of Jia et al. [42] to coarsely detect ellipse parameters (centers and axes). All the settings of the CNED were set as default except that the characteristic number on collinear points (CNL), the parameter of rejecting linearity, was set to 10.0 instead of 3.0. Due to thickness of ellipse edges on the wall, redundant ellipses could be detected using the CNED. The next steps averaged the center and axes' parameters over all redundant ellipses; edges were thinned using a morphological operator; and a finely fit ellipse [43] within each ellipse region was defined by its axes' parameters. The ID number within each ellipse region was identified by optical character recognition (OCR) in MATLAB, with the character set constrained to numbers 0–9. Using ellipse IDs greatly expedited the search for correspondence between x_t and P' from 70 camera frames. The accuracy of OCR recognition using MATLAB was about 80%. Improper IDs were later corrected by voting from the nearest four IDs, and the final recognition accuracy was 100% among the 70 frames. Given the corresponding x_t and P', the least-square Newton–Raphson iterative method [44–46] was applied to minimize residuals in Equations (2)–(6) following the aforementioned steps in Section 2.2. At the end of each iteration, the robust Huber's function [47] was applied to reduce the effect of residual outliers. Initial estimates of parameters were also required by the nonlinear Newton–Raphson method, in which α_0, β_0, and γ_0

were roughly set as 0°, 40°, and 0°; X_{c0}, Y_{c0}, and Z_{c0} were manually measured as 1.80, 1.08, and 1.09 in meters; and T, D_b, θ_0, ψ_0, ψ_Δ, ϕ_α, ϕ_β, ϕ_x, and ϕ_z were 0.03 m, −0.44 m, 180°, 90°, 2.5°, 0°, 0°, 0.06 m, and 0.00 m.

Figure 4. Experiment setup for the FSS calibration.

2.4. Fusion-Based Dense Point Cloud Recovery from FSS

Calibration is a preliminary requirement for fulfilling the framework in Figure 3. The framework focuses on generating and optimizing point clouds from multiple field scans. This section describes the suite of the visual odometry algorithms adopted, with a field experiment presented in the following section. Our field scans cover a hemispherical view with four scanlines spanning a vertical angle between 0° and 120°, though more scanlines can be added if desired. To avoid data overhead, we chose systematic sampling of the moving frames. Hence, our full hemispherical scan contains 150 × 4 moving frames with horizontal overlap of >90% and vertical overlap of ~80%. Aligning frames is the first problem, since the extrinsic parameters in the lab calibration environment are not repeatable in a new location. The intrinsic parameters, such as T, D_b, θ_0, ψ_0, and ψ_Δ from the Leddar sensor remain unchanged. To estimate the extrinsic parameters, target-based calibration or ground control points, while possible in some cases, would be tedious for a hemispherical view of 600 frames. Therefore, we directly used the dense and detailed photogrammetric information from the video to approximate camera poses R_t and then incorporated Leddar distance into the bundle adjustment for fine camera extrinsic parameters (R_t and $\boldsymbol{T}_t$).

A common way of aligning multiple frames is to extract invariant features in each frame, match features between frames, and optimize the camera colinear equation (Equation (6)). Frames with multiple scanlines also need iterative correction of scanline skewness caused by the uneven distribution of matching features. The set of global alignments has been supported by image stitching software, such as PtGUI, for this study. A weakness of using off-the-shelf stitching software is the limited number of identified matched pixels, which is insufficient for bundle adjustment needs. Therefore, intensive

extraction of SURF features [48] is added to the workflow in Figure 3, with outlier features filtered out using the homograph-based RANSAC algorithm [40]. The extracted SURF features are then matched between frames. Note that a 3D point can correspond to a set of matched pixels from multiple frames. The matched pixels from more than three frames are called "key pixels" here, whose features can be considered stable and will be used for the bundle adjustment later.

The global alignment uses PtGUI exports' Euler angles of each frame, which can be used to interpolate Euler angles (α_t, β_t, and γ_t) or R_t at any timepoint when Leddar distances are measured. Then Leddar point clouds $\boldsymbol{P}_t$ can be roughly recovered using Equations (2) and (3), assuming $\boldsymbol{T}_t$ is a zero vector. Projecting the Leddar point clouds $\boldsymbol{P}_t$ back to each image frame will add depth information to a few pixels (or "3D" pixels here). The depth values of the "3D pixels" are essential to the inference the depth values of the "key pixels." Our inference method is region-based interpolation: (1) filtering the foreground in each image using the k-means clustering method (k = 2); (2) segmenting images using statistical region merging algorithm (level = 8) [49]; and (3) implementing inverse distance weighting (IDW) interpolation for each key pixel within each region. Compared to global interpolation, using region-based interpolation better maintains sharp boundary lines between different image regions [50].

Based on the 3D "key pixels," both transformation matrices R_t and $\boldsymbol{T}_t$, and WCS coordinates $\boldsymbol{P}_t$ can be finely estimated using iterative bundle adjustment. First, $\boldsymbol{P}_t$ estimated from 3D "key pixels" are averaged among different frames and reprojected to each image frame using Equation (7):

$$\begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = KR_t^{\mathrm{T}}\left[\begin{pmatrix} X_t \\ Y_t \\ Z_t \end{pmatrix} - \boldsymbol{T}_t\right] \tag{7}$$

$$\min_{\alpha_t,\beta_t,\gamma_t,X_{ct},Y_{ct},Z_{ct}}\left\|\begin{pmatrix} x'_t \\ y'_t \end{pmatrix} - \begin{pmatrix} x_t/z_t \\ y_t/z_t \end{pmatrix}\right\|, \text{ using nonlinear regression} \tag{8}$$

$$\begin{pmatrix} X'_t \\ Y'_t \\ Z'_t \end{pmatrix} = R_t K^{-1}\begin{pmatrix} x'_t \\ y'_t \\ 1 \end{pmatrix} + \boldsymbol{T}_t \tag{9}$$

$$\min_{X_t,Y_t,Z_t}\left\|\begin{pmatrix} X_t \\ Y_t \\ Z_t \end{pmatrix} - \begin{pmatrix} X'_t \\ Y'_t \\ Z'_t \end{pmatrix}\right\| \text{ using ridge regression } (\lambda = 0.01) \tag{10}$$

The reprojected points are normalized by Z coordinates and compared to the 2D coordinates of the "key pixels." The least square error is minimized using nonlinear regression, with the camera's extrinsic parameters (α_t, β_t, γ_t, X_{ct}, Y_{ct}, and Z_{ct}) as variables (Equation (8)). The new $\boldsymbol{P}'_t$ corresponding to "key pixels" is estimated using Equation (9) based on the optimized camera extrinsics. Note that $\boldsymbol{P}_t$ is sensitive to small errors of "key pixels" and camera extrinsics due to the ill-posed mono-camera geometry. A robust solution to the ill-posed optimization problem is to use ridge regression [51], which partially minimizes the least square error between $\boldsymbol{P}_t$ and $\boldsymbol{P}'_t$ in Equation (10). Its regularization parameter λ is set to be 0.01. The optimized $\boldsymbol{P}_t$ from Equation (10) is again reprojected to each image frame in Equation (7), and repeat iterations from Equation (7) to Equation (10) until the error in Equation (8) is locally minimal. This iterative bundle adjustment for 3D recovery is similar to the gold standard method from Hartley and Zisserman [40], but with depth information provided from sparse Leddar distance instead of dense stereo geometry. The resulting point clouds will exhibit rich details in the 2D planar direction but limited variation in the depth direction.

The 3D recovery from bundle adjustment produces point clouds for "key pixels." The "key pixels" are essentially from SURF feature extraction and are mostly focused on corner pixels with sharp color gradients. Other "internal" foreground pixels should also be incorporated to produce complete and dense point clouds. Our method is to extract dense foreground pixels and solve Equations (9) and (10) to create optimal dense 3D points. To satisfy Equation (10), one 3D point requires at least one pair

of matching pixels from two frames. We already have matching pixels defined as "key pixels" from previous steps. We need to extrapolate the matching relationship for all foreground pixels. This is a pixel-level dense matching process. First, the foreground pixels need to be subsampled at a certain interval (e.g., 10 pixels in this study) to avoid data overhead. Then the disparities of "key pixels" between current frame and one matched frame are calculated. The disparities are used to interpolate a disparity map for all foreground pixels in the current frame. Given a disparity map, the foreground pixel location in the matched frame can be estimated. This dense matching step between a pair of matched frames is repeated for all the matching frames. Each set of matched pixels from multiple frames is given one unique ID, corresponding to one unique P_t. Finally, using Equations (9) and (10), the densely matched sets of pixels can produce dense point clouds. The final RGB colors of dense point clouds are the average RGBs from matched pixels.

2.5. Application: Tracking the Autumn Leaf Drop Processes with the FSS

The FSS was used to track canopy changes during an autumn season in 2018. Our experiment's site was located in an area of cottonwood (poplar) stands in Lethbridge, Canada (49°41′45.2′′ N, 112°51′54.0′′ W). The FSS was mounted on a tripod surrounded by six poplar trees within 20 m, including *Populus angustifolia*, *P. deltoides* and their hybrid, *P. × acuminata* [52]. The irregular shapes of poplar trees increased the difficulty of depth information recovery but the rich texture of the scene facilitated feature extraction, which supports frame alignment. The camera lens filter was replaced to block near-infrared light and enable natural-colored images. Multiple scanlines were collected, each corresponding to 360° horizontal rotation at a speed of 2.4° per second. Four scanlines were selected to cover the upper hemispherical view of the scene for further processing. All scans were aligned, optimized, and densified to create colored point clouds using the methods in Section 2.4. The same scene was scanned with a Teledyne Optech ILRIS HD (1535 nm) TLS as a benchmark. The scanning angle was 360° × 80°, with the small zenith angle between 0°–10° not scanned. Only last returns were recorded and point spacing was 3.2 cm at a distance of 20 m from the TLS. A total of 30 ILRIS TLS scans were collected in 30 minutes, with three scanlines covering the entire upper hemispherical view. The TLS scans were co-registered by the iterative closest point (ICP) algorithm into one hemispherical scan with an average accuracy below 1.3 cm. The same scanning and processing activities were repeated on September 9th, September 17th, October 1st, and October 17th during the autumn defoliation period in 2018, to evaluate the reusability of our static scanning system in a temporal monitoring context. Hemispherical photos based the digital hemispherical photography (DHP) methods were also captured on September 9th, September 17th, and October 17th for benchmarking purposes.

Canopy vertical volume profile and plant area index (PAI) were extracted from the FSS point clouds to evaluate the capabilities of 3D canopy detection and canopy attribution. The volume profile was defined as a total volume of voxels at each height, where a unit voxel was 0.1 × 0.1 × 0.1 m and a height slice was 0.1 m. The PAI was calculated based on a path length distribution (PATH) model [7,15]. Specifically, the PATH model consisted of two equations (Equations (11) and (12)) with the gap fraction $\overline{P(\theta)}$ and path distribution p_l as the only two inputs. To calculate angle-specific gap fractions, a hemispherical image from either FSS or TLS point clouds was first converted to a black and white binary image under a fisheye perspective. The fisheye binary image was equally sliced into 28 rings representing a zenith angle between 15°–69° and a ring width of 4°. The overlap between two neighboring rings was 2°. The gap fraction $\overline{P(\theta)}$ was defined as the ratio of the "hole" pixel numbers to the "filled" pixel numbers within a ring slice. The "filled" pixels represented the overall canopy area and was generated based on image morphological smoothing. The path distribution p_l was defined as the probability density function (PDF) of the optical path length within the crown area, with l representing the within-crown path length. The l ranged between 0 and 1, scaled by the maximum value l_{max}. The p_l was approximated by the histogram of the l, normalized by the total histogram area. With both the gap fraction $\overline{P(\theta)}$ and path distribution p_l extracted from crown area, the integral equation (Equation (11)) was solvable based on any root-finding algorithm, and

the $FAVD \cdot l_{max}$ could be estimated; $PAVD$ stands for the plant area volume density and $G(\theta)$ the leaf angle distribution. The $G(\theta)$ was set to 0.5 in this study, corresponding to a spherical leaf angle distribution [53]. The $PAVD \cdot l_{max}$ was then input to Equation (12) to determine the $PAI_{true}(\theta)$ at a specific zenith angle θ. The final PAI value was a weighted sum of $PAI_{true}(\theta)$ over all zenith angles (Equation (13)) [15].

$$\overline{P(\theta)} = \int_0^1 e^{-G(\theta) \cdot (FAVD \cdot l_{max}) \cdot l} p_l dl, \text{ where } \int_0^1 p_l dl = 1 \tag{11}$$

$$PAI_{true}(\theta) = \int_0^1 \cos(\theta) \cdot (FAVD \cdot l_{max}) \cdot l \cdot p_l dl \tag{12}$$

$$PAI_{true} = \frac{\sum_\theta PAI_{true}(\theta) \cdot \sin(\theta)}{\sum_\theta \sin(\theta)} \tag{13}$$

An important portion of the canopy was foliage and the corresponding index was LAI. The LAI was directly related to photosynthetic processes and carbon productivity, and was a more sensitive index than PAI to reflect seasonal biomass changes. We estimated LAI values by contrasting leaf-on and leaf-off gap fraction values, as illustrated in Equation (14), where N was the number of pixels and P was short for the gap fraction $\overline{P(\theta)}$. Specifically, N_{wood}, N_{leaf}, and N_{hull} are the numbers of wood, leaf, and canopy pixels, respectively, and P_{off}, P_{on}, and P_{leaf} were the gap fractions of leaf-off, leaf-on, and leaf-only canopies, respectively. Based on Equation (14), P_{leaf} was a simple ratio of P_{off} to P_{on} (Equation (15)). With known P_{leaf}, LAI was estimated in a similar manner with PAI using Equations (11)–(13), except for replacing $\overline{P(\theta)}$ (namely, P_{on}) with P_{leaf}.

$$P_{off} = 1 - \frac{N_{wood}}{N_{hull}}, \; P_{on} = 1 - \frac{N_{wood} + N_{leaf}}{N_{hull}}, \; P_{leaf} = 1 - \frac{N_{leaf}}{N_{hull} - N_{wood}} \tag{14}$$

$$P_{leaf} = \frac{P_{on}}{P_{off}} \tag{15}$$

## 3. Results and Discussion

3.1. Calibration

An example frame among the 70 frames from the camera video is shown in Figure 5a. The original 720p frame is cropped to 1280×500 to be compact. Only part of the 16×9 circle grid is within the field of view. The frame shows no blurry effect, implying that an FPS of 25 is sufficient to match a rotation speed of 2.4° per second. The purple band in the frame is the near-infrared LED light from Leddar, because the camera lens has no IR filter. The visibility of the LED light provides an intuitive way of validating calibration accuracy: the footprints of all the 16 Leddar segments after calibration should fall within the purple area. The ellipses on the wall were detected with the CNED method, shown in green in Figure 5b. It is clear that a few incomplete ellipses were skipped and many redundant ellipses were created. That is a preliminary step of approximating ellipse ranges and locations. Fine ellipses after edge thinning and geometrical fitting are shown in random colors in Figure 5c overlaid by the edge image. Edge noise is inevitable but has a limited impact on the ellipse fitting results. Each ellipse center is marked as a green cross. The integer number inside each ellipse is the circle ID predicted with the OCR and posterior voting method. The OCR recognition confidence is also placed under each ellipse ID as a decimal number. Ellipses with low OCR confidence are removed, such as numbers 142 and 80. The remaining 33 ellipses still satisfy the minimal requirement of having four control points for the camera collinearity equation. Note that all the 70 frames have been inspected to have four or more ellipses at the beginning. After calibration, point-based footprints of the Leddar segments were

projected on the example frame. The Leddar points basically fall within the LED light zone, except for the first segment on the less illuminated area.
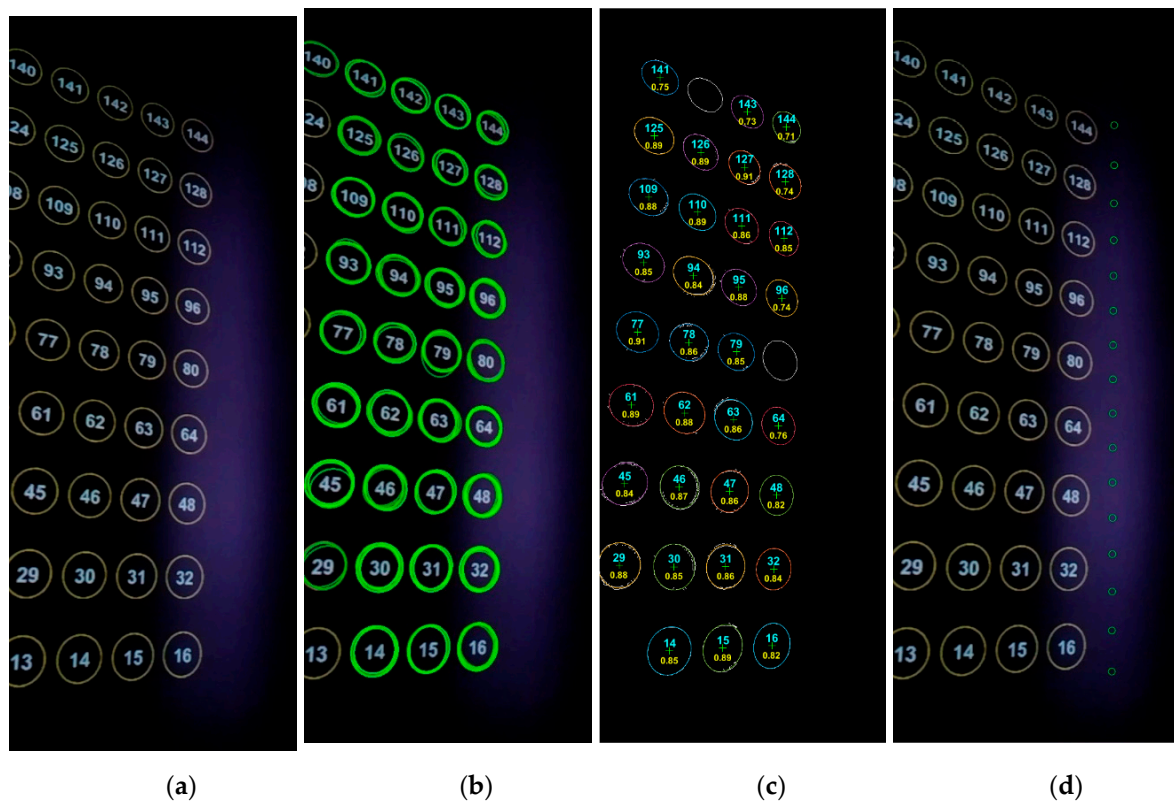
Figure 5. Calibration processing of an example frame: (**a**) circle grid (yellow) and LED light (purple) from camera view, (**b**) ellipse detection (green) by CNED method, (**c**) fine ellipse fitting (random color), ellipse ID from OCR (cyan) and OCR confidence (yellow), and (**d**) calibrated segment points reprojected to the frame image (green).

The point clouds after calibration of Leddar distances and poses are displayed in Figure 6a, with the X–Y projection approximating a planar shape and the X–Z projection a linear shape. The entirety of the point clouds contains 16 segments, each creating 70 points along the horizontal rotation direction. Several points overlap when the Leddar is still static at the beginning or the end. The projected trajectory of each segment on the X–Y plane is not a straight line, because Leddar has a fixed tilt angle of approximately 15° upwards. A few lines are not smooth, and their noise is not systematic for all lines, probably less likely due to the servo movement but rather the Leddar distance's measurement instability. The standard error is 1.03 pixels for optimizing the camera collinearity equation (Equation (6)), 3.84 pixels for optimizing the temporal rotation equation (Equations (4) and (6)), and finally, 9.7 mm in WCS for solving the Leddar distance equation (Equations (2), (3), and (5)). Equation (4) constrains the rotation matrix to a fixed rotational axis. Without Equation (4), solving Equation (6) for each frame separately is still feasible and yields a standard error of 2.34 pixels. However, the retrieved camera extrinsic parameters, such as the camera center locations T_t, shown as the green dots in Figure 6b, lose physical meaning and present irregular movements. Therefore, applying Equation (4) accounts for the real camera movement, shown as the white arc points in Figure 6b, thus enables more reliable calibration parameters. The final estimates of intrinsic parameters T, D_b, θ_0, ψ_0, ψ_Δ, ϕ_α, ϕ_β, ϕ_x, ϕ_z are 0.070 m, −0.428 m, 180.84°, 89.39°, 3.32°, −0.69°, 12.19°, 0.027 m, and −0.0006 m.

(**a**)   (**b**)
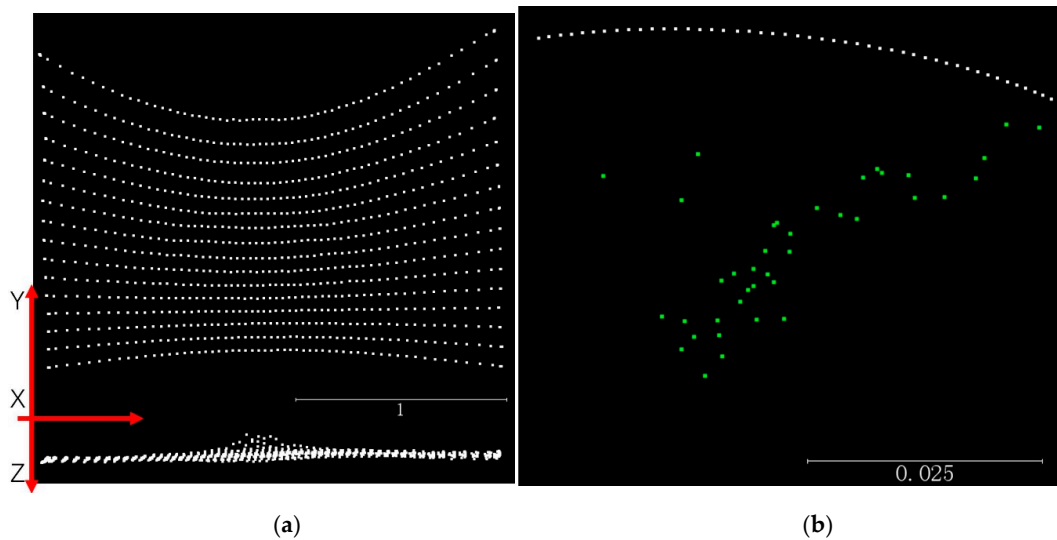
Figure 6. Calibrated Leddar points and camera trajectory: (**a**) calibrated Leddar points of a flat wall on the X–Y plane (above) and on X–Z plane (below), and (**b**) camera trajectory with rotational constraints (above) and without constraints (green below).

3.2. Fusion-Based Dense 3D Recovery

The success of point cloud recovery hinges, to a significant degree, on the quality of aligning video frames, because the camera poses determine the general form and structure of point clouds. Our video scans of poplar trees on four different dates are aligned based on automatic solutions provided from PtGUI, including image matching, feature extraction, feature matching, horizon correction, bundle adjustment, and image mosaicking. Example alignment results for the October 1st and October 17th videos are visualized as 360° × 120° spherical panoramas under equirectangular projection in Figure 7a,b. The trees displayed leaf-on conditions on October 1st and were defoliated completely on October 17th. Both scenes were centered on a railway viaduct, and the lower part of the panorama was discarded. No obvious alignment gap or inconsistencies were found from the two images. The processing results of the leaf-on scene are visualized in Figure 7c–h. Figure 7c shows alignment of separate images and their seamlines in PtGUI without mosaicking and color blending. The colors of individual tiles in Figure 7c differ from each other due to sunlight variation during scanning. Yet based on visual inspection, the alignment of tiles is seldom affected by the color difference, indicating strong robustness of PtGUI's feature extraction algorithms. The alignment errors estimated from bundle adjustment in PtGUI are 4.5, 3.0, 2.5, 2.5, and 2.2 pixels for the scenes of September 9th, September 17th, October 1st, and October 17th, respectively. The relatively large error of the September 9th is due to various factors, such as windy conditions, the thick canopy, and a cloudy sky.
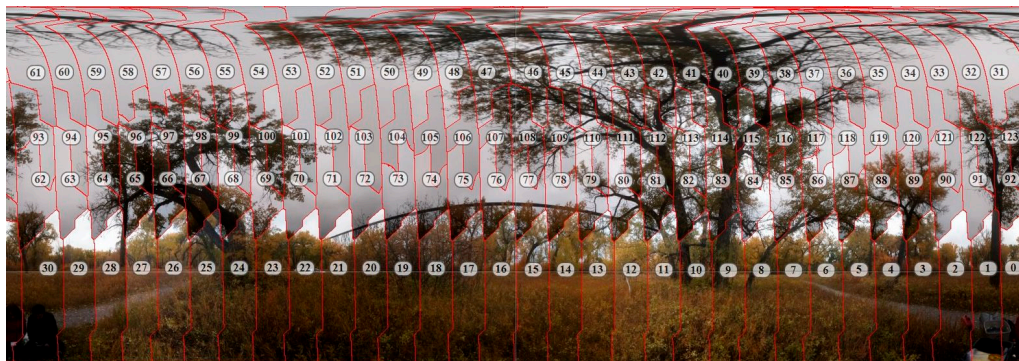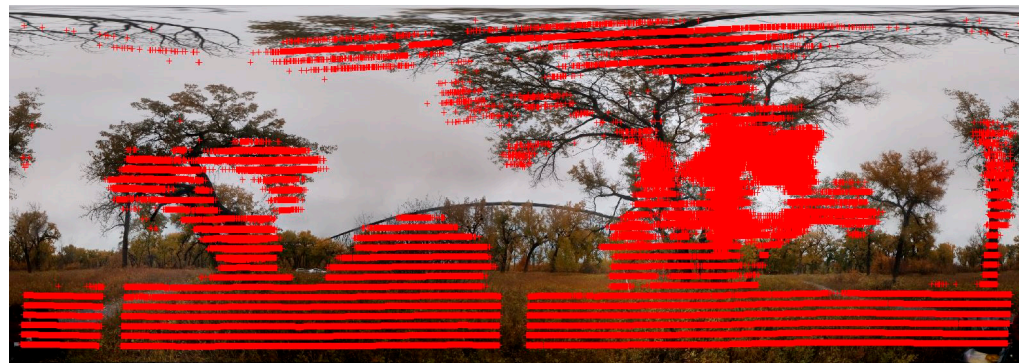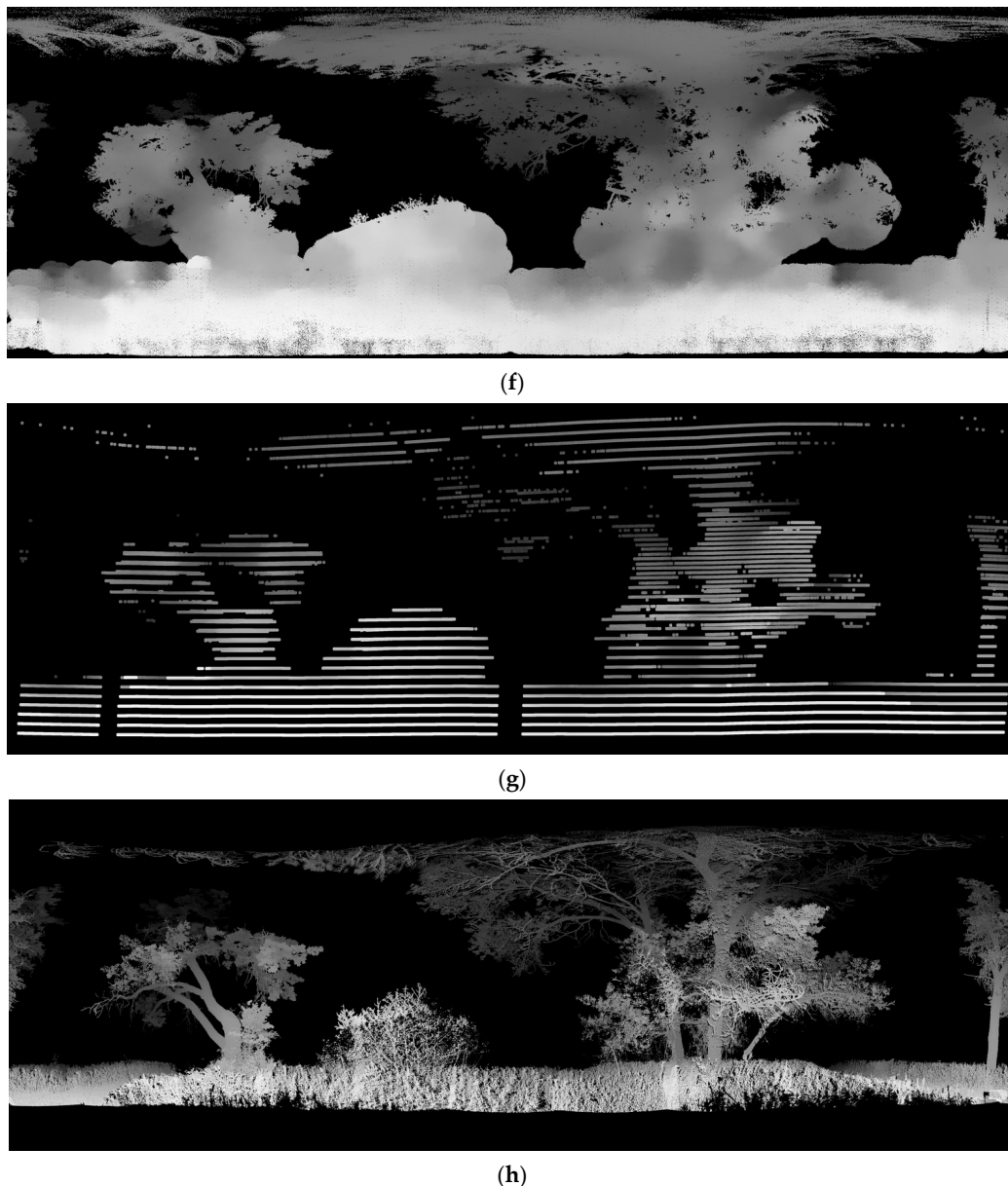

(**a**)

Figure 7. *Cont.*
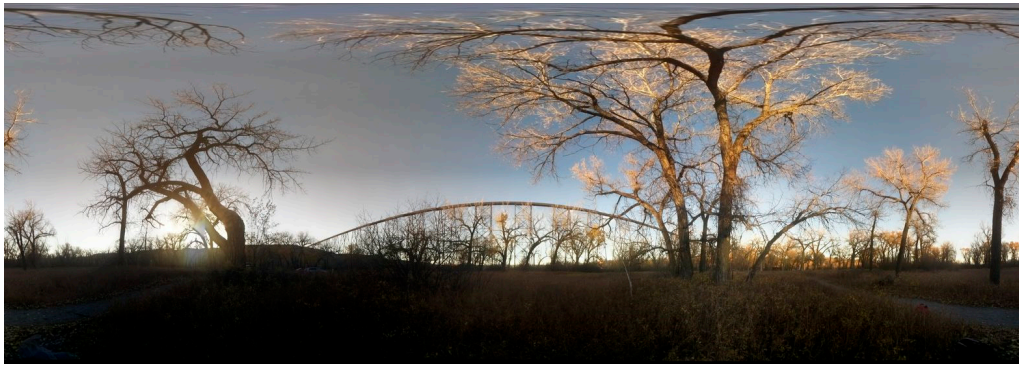
(**b**)

(**c**)

(**d**)

(**e**)

Figure 7. *Cont.*

(**f**)

(**g**)

(**h**)

Figure 7. Hemispherical view of processing results: (**a**) image global alignment for the October 1st scan, (**b**) image global alignment for the October 17th scan, (**c**) global alignment layout in PtGUI software with image IDs and seamlines for the October 1st scan, (**d**) Leddar-only point clouds (red cross) reprojected to the hemispherical image, (**e**) RGB colors from fusion-based point clouds, (**f**) depth image from fusion-based RGB point clouds, (**g**) depth image from Leddar-only point clouds (point size enlarged for clearer visualization), and (**h**) the depth image from TLS scans. (**e**–**h**) were all done using hemispherical projection.

Leddar points were reprojected as the red crosses in Figure 7d and overlaid with the panorama view of the October 1st scene, after applying Leddar intrinsic parameters from the calibration, and camera extrinsic parameters from PtGUI alignment. The Leddar points capture the basic scene structure nearby, except for upper canopy, distant ground, and thin branches. The minimum, average, and maximum detection ranges of Leddar in this scene are 1.64, 6.45, and 14.17 m, respectively. The Leddar point clouds have obvious gaps between the segments and on the ground due to missing signals. This problem of Leddar data sparsity limits potential applications, such as tree surveying and object detection, unless photographic information is integrated. Therefore, the iterative bundle adjustment

is applied at the point cloud level to minimize the disagreement between Leddar reprojected pixels, camera pixels, and camera extrinsic parameters. Iterations of the bundle adjustment error, measured in pixels, are plotted for the four scanning dates in Figure 8. The initial error of bundle adjustment can be more than 8 pixels but will converge to a level comparable to PtGUI alignment error. The final errors from bundle adjustment were 2.8, 1.9, 2.0, and 1.8 pixels for September 9th, September 17th, October 1st, and October 17th, respectively.
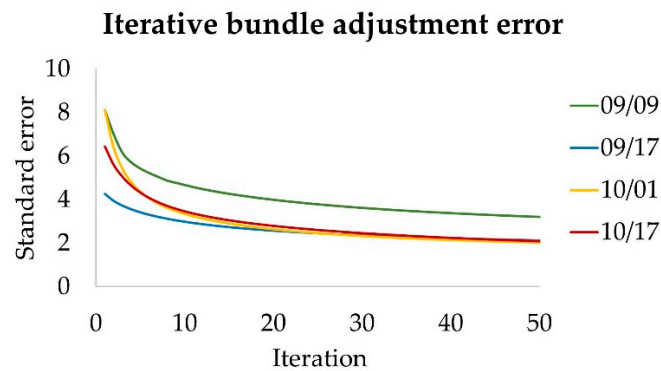
Figure 8. Fusion error convergence with iterations (in pixels) on four defoliating dates in 2018.

The fusion-based point clouds after image background removal, iterative bundle adjustment, and dense matching recovery were reprojected into two panorama images, as shown in Figure 7e–f. Figure 7e displays reprojected pixels with RGB colors, and Figure 7f is the corresponding depth image with nearer objects showing brighter colors. The reprojection from point clouds to a hemispherical-view image is not simply one point per pixel, considering the previous dense recovery process has a subsampling rate of 10 pixels per point. Therefore, each point had a buffer of 10 pixels in a hemispherical image. Similarly, reprojecting TLS point clouds into a hemispherical depth image in Figure 7h also needs to consider the footprint of each TLS laser beam. The scanning spacing of each Ilris HD beam (1600 μrad) was set to be the footprint size according to the suggestions in [12]. The scanning spacing corresponds to a constant pixel size of 1.6, thus each reprojected pixel of the hemispherical images was dilated by a factor of 1.6.

In contrast to the Leddar reprojection image in Figure 7d, the image in Figure 7e not only captures rich 2D details but also covers a reasonable extent due to the region-based interpolation. The main problem of the fusion-based point clouds is false interpolation. The problem can be illustrated when comparing specific tree point clouds extracted from TLS, Leddar point clouds, and fusion-based point clouds in Figure 9a–c. The TLS point clouds clearly exhibit branch-level details with warmer colors representing higher laser intensity. The fusion-based point clouds have distinguishable stem colors and noisy branches, still highly detailed compared to the obscure Leddar point clouds. Yet the fusion-based point clouds overfill the gaps between branches and also falsely incorporate pixels from remote shrubs. This is inevitable because region-based interpolation and bundle adjustment can mitigate but not eradicate the problem of coarse and sparse depth measurement from Leddar. The depth image in Figure 7f displays a strong smoothing effect compared to the depth image reprojected from TLS in Figure 7h, but is much more detailed than the Leddar-only point clouds in Figure 7g with indiscernible sparse points.
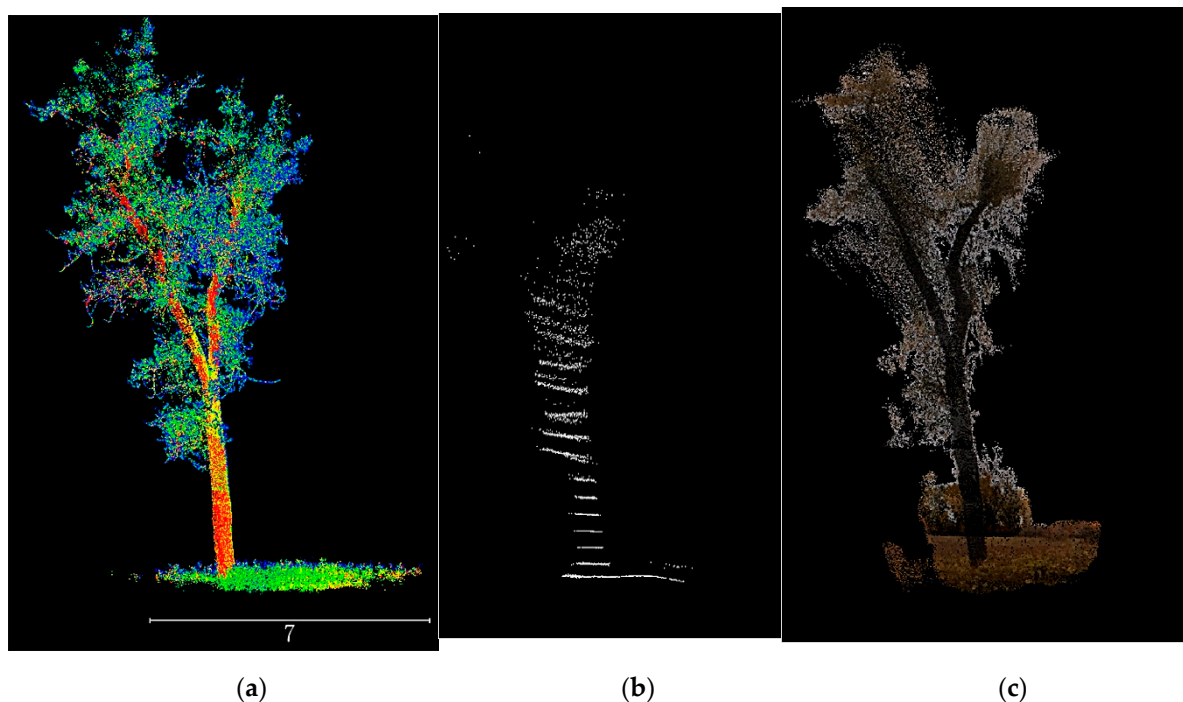
Figure 9. Example tree point clouds from (**a**) TLS scans, (**b**) Leddar-only point clouds, and (**c**) fusion-based dense point clouds.

3.3. Tracking Changes of Canopy Vertical Volume Profile, PAI, and LAI

Vertical volume profiles from TLS, fusion-based point clouds, and Leddar point clouds can be compared in Figure 10a–d. The r^2 of profiles over the maximum height range between Leddar and TLS was noted for each date. The *p*-value based on a paired t-test between profiles is also provided. It is a rule of thumb to consider that two profiles have different mean values if the *p*-value is below a significance level of 0.05. Regardless of scanning date, both the profiles from the fusion-based point clouds and the Leddar point clouds were correlated with the TLS profiles. The r^2 between the Leddar and TLS profiles remains at approximately 0.3 from the first three leaf-on scenes and increases to 0.48 for the last leaf-off scene. In contrast, the r^2 of profiles between the fusion-based and TLS are around 0.65 from the leaf-on scenes, constantly higher than 0.52 from the leaf-off scene. The r^2 improvement of fusion-based over Leddar is because thick crowns and leafy understory lead to Leddar signal loss but do not affect photography-based interpolation. The profile difference between the fusion-based and the reference TLS is mainly the higher frequency in the middle crown area due to the overfilling effect, and also thinner volume near the upper crown associated with the loss of supporting points from Leddar. For the first two leaf-on scenes, the overestimation effect near the middle crown is dominant, thus causing an obvious bias of mean values denoted by a low *p*-value of 0.000. For the last two scenes with increasingly defoliated crowns, the fusion-based point clouds tend to incorporate fewer false pixels from areas beyond the canopy, resulting in the retreat of the lower canopy. The fusion-based point clouds' points are not as rich in the depth direction compared to TLS, thus the lower canopy parts of the last two scenes are thinner than TLS. As a result, the mean bias of profile is offset, and high *p*-values (>0.3) are found for the last two scenes. Note that the TLS has a slightly narrower scanning view than the FSS, with part of the upper crown and ground not sampled by scans. The profile difference around the upper crown and ground can be higher than observed in Figure 10. This problem of profile distortion might be due to the imperfect hemispherical stitching process.
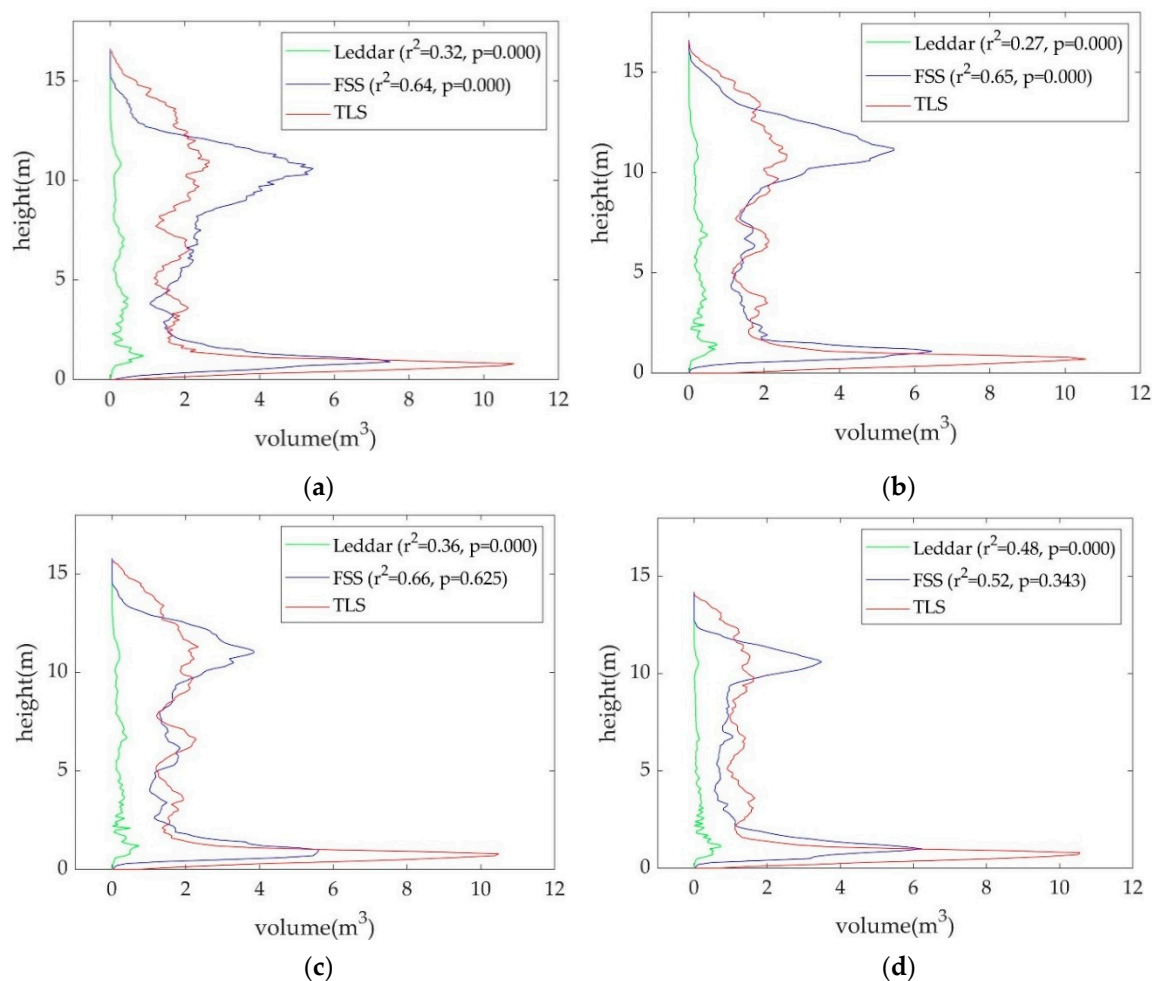
Figure 10. Vertical volume profiles from TLS, fusion-based, and Leddar-only point clouds on (**a**) September 9th, (**b**) September 17th, (**c**) October 1st, and (**d**) October 17th, 2018, where the horizontal axis denotes volume of voxels with a unit voxel of 0.1 m^3, and the vertical axis denotes height in meters.

The benefits of synthesizing both 3D and color information make FSS a potentially valuable complement to conventional LAI or PAI surveying tools, such as digital hemispherical photography (DHP). Figure 11 compares the fisheye image from FSS with the DHP photo from the same site. The canopy shapes in the two images are visually identical, seen in Figure 11a,b. The FSS, in addition, captures depth information shown in the image in Figure 11c, with a benchmarking TLS depth image provided in Figure 11d. Note that the upper crown area was not scanned with the TLS due to the field of view constraint. The availability of depth images enables FSS to calculate true PAI and LAI based on the PATH model. Different PAI and LAI estimates based on FSS, TLS, and DHP methods, and based on PATH and non-PATH methods, are contrasted in Figure 12, with bars denoting PAI and crosses indicating LAI. The non-PATH method relies on the LAI models from the Hemisfer software [54,55], which combines the leaf angle distribution (LAD) function of Lang [56]; the clumping correction of Lang and Xiang [57]; and the non-linearity correction model of Schleppi, Conedera, Sedivy, and Thimonier [54]. The non-PATH method focuses on the RGB images from DHP or FSS, or the depth images from TLS, whereas the PATH method additionally needs point cloud input from FSS or TLS.
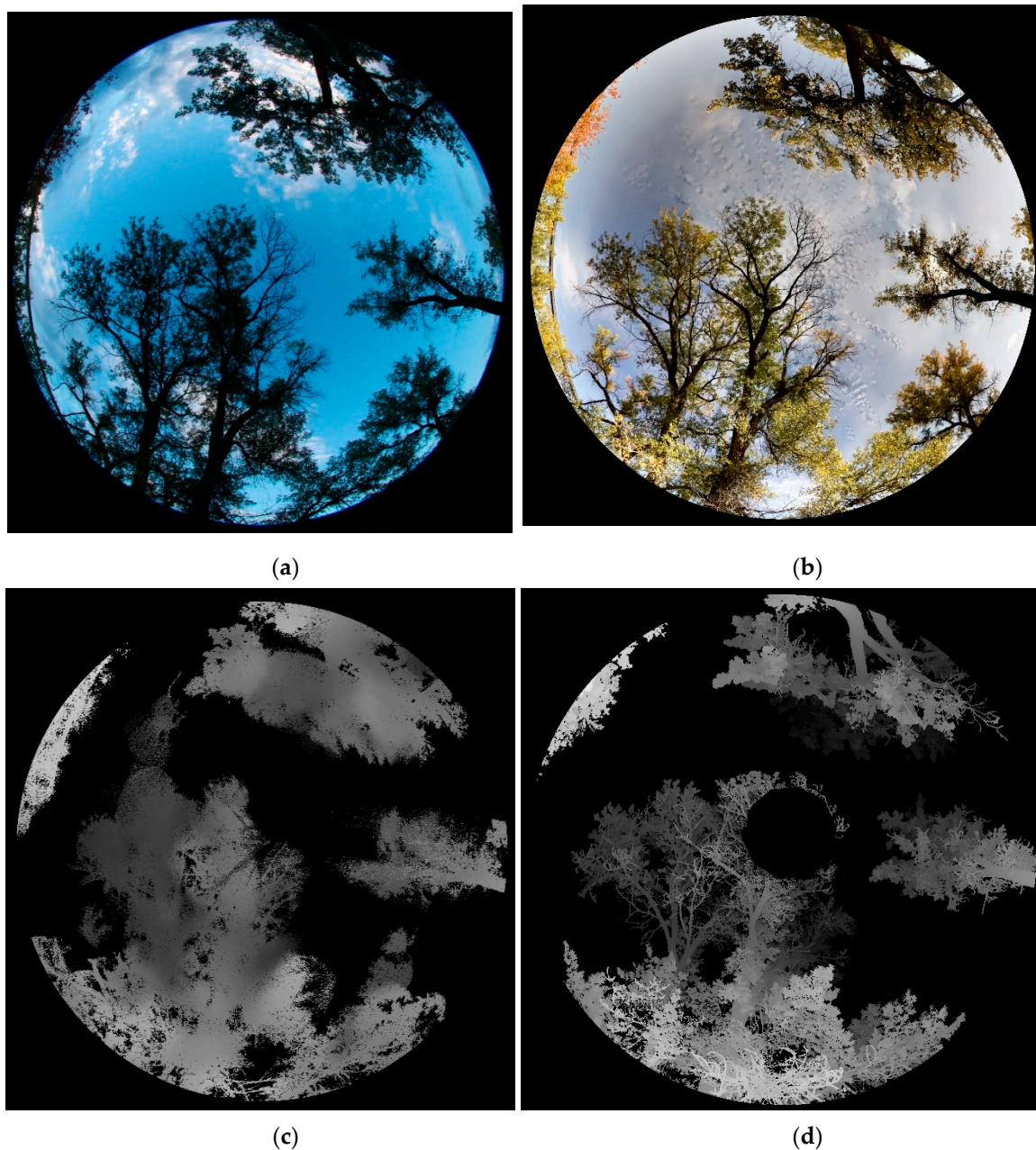
Figure 11. Fisheye-view images compiled from the September 9th datasets based on (**a**) DHP, (**b**) FSS, (**c**) FSS depth, and (**d**) TLS depth.

For the non-PATH methods in Figure 12, PAI and LAI values generally decline with the defoliation dates, with all the LAI values reaching zero level on the leaf-off date October 10th, except that the PAI and LAI value from non-PATH FSS increases on October 1st. The incorrect increase implies the instability of using image-only methods. The possible cause of the incorrect increase is FSS's underestimation of PAI and LAI from the September 9th and October 1st FSS images, in contrast to the PAI and LAI values from DHP and TLS. Strong spectral reflectance from sunlight is observed in the September 9th and October 1st FSS images and a small portion of canopy pixels in the FSS image displays a similar color as the sky background. These bright canopy pixels were not successfully identified as leaf area, causing the underestimation effect. The DHP method does not have the underestimation issue because the DHP images were captured near dusk. The TLS method does not have the stability issue of FSS, due to the fact that depth images are used instead of color images. The TLS method, however, has an issue of overestimating PAI. The leaf-off PAI from TLS is 32% higher than from DHP, compared to the

average 3% overestimation of leaf-on PAI from TLS. The overestimation issue of TLS has two typical causes. The depth images, particularly the leaf-off ones, contain ghost points or misaligned points around thin branches and twigs. Gaps smaller than the beam width of TLS are also not differentiable from the depth images [12]. The overestimation by TLS of twig PAI leads to an underestimation of LAI by 26% relative to DHP.
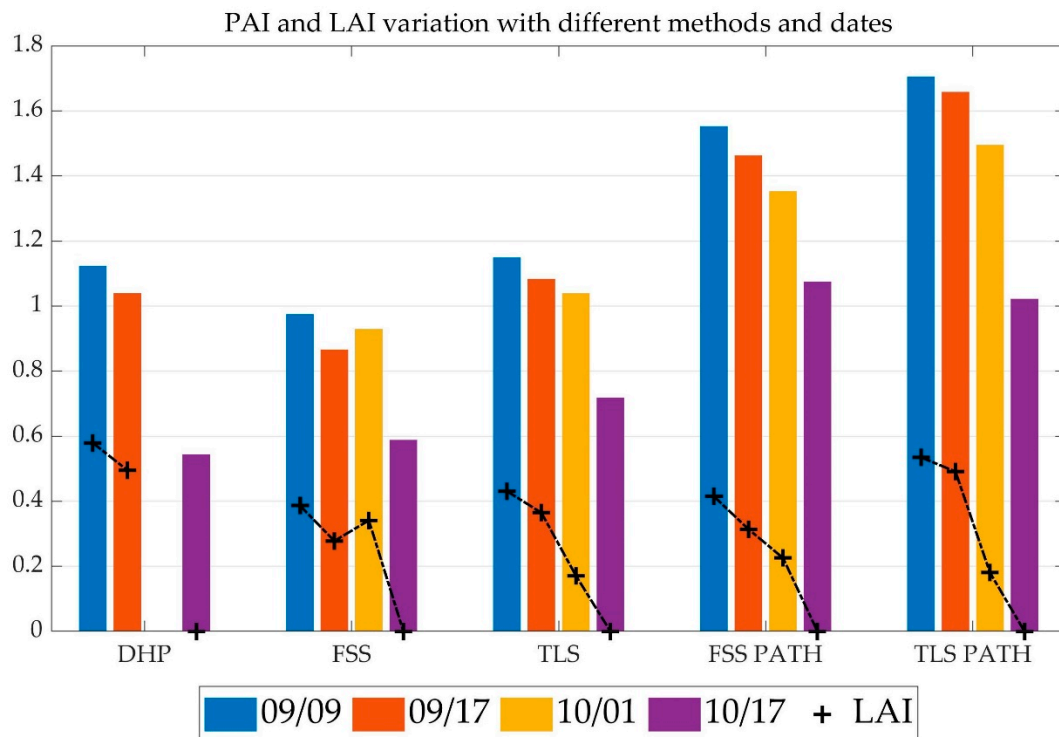

Figure 12. Comparing different methods of leaf area index (LAI) estimation on four scanning dates. The colored bars represent plant area index (PAI), and crosses for the associated LAI. The October 1st DHP dataset is not available. The DHP and the FSS methods are based on RGB images and the TLS based on depth images.

With the PATH model applies to TLS and FSS, the PAI estimates are approximately 30%–45% higher than the non-PATH PAIs. The PATH PAI estimation from FSS does not have the problem of PAI increase on October 1st, indicating the importance of incorporating depth correction. The PATH LAI estimates are also higher than the non-PATH by 14% on average, except for the October 1st FSS LAI anomaly. Considering that the optical image methods usually underestimate true PAI or LAI [7] by 20%–60%, it is assumed that the PATH model is a closer approximation of the true PAI or LAI values.

It is important to understand why the PATH model usually outputs higher PAI (or LAI) values than the classic geometrical-optical model. Indeed, the $PAI_{true}(\theta)$ solved from the PATH model does not have a simple analytic form, due to various forms of p_l. However, if we simply assume that p_l is constantly 1, or equivalently, the within-crown path length distribution is uniform, the PATH model then has an analytical solution of PAI, which is basically a Lambert W function of gap fraction $\overline{P(\theta)}$ (Figure 13). The traditional LAI model (effective LAI) using Beer's law [14,58] is also contrasted in Figure 13. It clearly shows that the PATH PAI is consistently higher than the non-PATH PAI, especially when the gap fraction is small. It is also noteworthy that the PATH model might be overly sensitive to near-zero gap fraction changes. The upper bound of PAI based on the PATH model is $-\frac{\cos(\theta)}{G(\theta)}\frac{l_{max}}{l_{min}}ln\overline{P(\theta)}$ and the lower bound is $-\frac{\cos(\theta)}{G(\theta)}\frac{l_{min}}{l_{max}}ln\overline{P(\theta)}$. The wide range of PAI indicates strong flexibility of the PATH model, but it is also important for future studies to examine what the rigorous PAI bounds based on different forms of the p_l functions are, and what a suitable analytical form of p_l functions

or PAI functions with a smoother gap fraction sensitivity can be. In addition, the PATH model is essentially a variant of Beer's law. It does not account for the effect of laser footprint increase and density attenuation with distance. Integrating these laser-dependent factors to the PATH model is feasible, but solving the integral equations would become difficult. An alternative approach provided by [59] is to discard Beer's law and model the statistical form, relating leaf area distribution, laser path length, density attenuation, and footprint size. Such a statistical model is solvable with a maximum likelihood estimator (MLE).
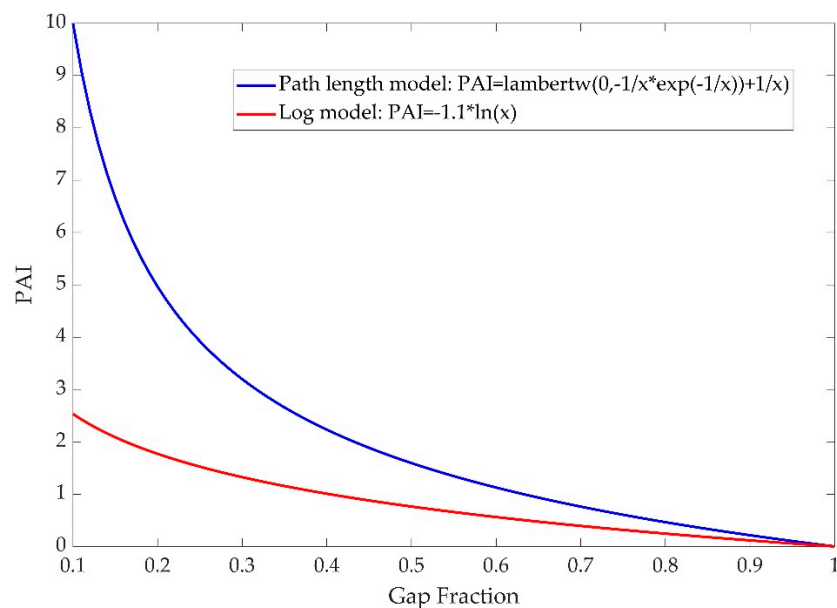
Figure 13. Relationship between gap fraction and PAI (or LAI). The red curve shows PAI values from the path length distribution model, compared to the blue curve from the simple Beer's law model. Detailed mathematic functions are provided in the legend, with *x* representing the gap fraction.

4. Conclusions and Future Work

Timely monitoring of canopy characteristics is necessary to understand the spatiotemporal variation of biomass in a forest ecosystem and to evaluate carbon budgets as part of forest stand reporting. The advent of low-cost multi-segment LiDAR sensors, Leddar in particular, has presented many successful object tracking applications. Yet the Leddar sensor is not comparable to TLS in sampling 3D details due to its limited FOV and point resolution. This limitation was mitigated in this study by constructing a low-cost 3D fusion scanning system, FSS, integrating Leddar, camera, and pan-tilt robotics. A framework of integration was developed, generally including (1) plane-based physical calibration, converting Leddar distance into a 3D point and locating the Leddar point from images; (2) global image alignment, obtaining panorama and coarse camera poses; (3) iterative bundle adjustment, optimizing camera poses using both Leddar distance and corresponding pixels at the point cloud level; and (4) dense point cloud recovery, based on density matching and interpolation. The calibration error of Leddar points was 9.7 mm at a distance of ~1 m. The set of fusion-based methods was applied to recover hemispherical colored point clouds from multi-temporal poplar tree scans during the autumn defoliation period. The bundle adjustment error was 1–3 pixels, indicating a strong agreement between image and Leddar projection on the X–Y plane. However, uncertainty existed in the depth (Z) direction due to the coarse resolution of Leddar distance. Final fusion-based point clouds were compared to the TLS scans collected on the same spot and date. The vertical profile volumes between TLS and FSS point clouds had an r^2 of 0.5–0.7 over the maximum tree height range, which varied with leaf cover conditions and exceeded the r^2 of 0.3–0.5 between TLS and pure Leddar point clouds. PAI and LAI metrics were also extracted from FSS, TLS, and DHP for leaf-on and -off dates.

Using only image data, the PAI and LAI tended to be underestimated with FSS and overestimated with TLS. With the point cloud PATH model, both PAI and LAI from FSS or TLS were corrected to approach their assumed true values. By combining both color and depth information, the FSS has demonstrated versatility and significant potential for the application of canopy foliage monitoring.

The FSS system was primarily developed for static scanning of the environment. For environmental applications, such as crown measurement and biomass delineation, low-cost sensor systems such as the FSS (as built) do not parallel the resolution or precision of TLS and DHP at present, yet the demand for gross mensuration should not be overlooked, and sensor hardware upgrading is inevitable. The core contribution of this study is a holistic calibrating and fusing scheme for a low-resolution multi-sensor platform. The advantage of FSS should be clear. It is more portable and of a lower cost compared to a conventional TLS, and has a higher frequency, and improved detection range and FOV compared to many indoor-oriented LED or flash LiDAR systems. It is, therefore, suitable to be deployed in high numbers into sensor networks for broad-scale environmental monitoring, which will be our future work. Adapting the FSS to other mobile systems such as UAV is also feasible, following a similar framework of calibration, pose approximation, bundle adjustment, and densification, except for the need for external pose information from GPS and IMU, and a dedicated image alignment method. A better viewing geometry, such as stereoscopy from a mobile system, could substantially improve the precision of 3D recovery, particularly in the depth direction. In addition, the fusion scheme presented in this study focuses on the data aggregation level, whereas a higher level of fusion based on spatiotemporal attributes, patterns, and mission management still requires investigation. A bright future for cost-effective, fusion-based 3D canopy monitoring systems is anticipated.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. van der Sande, M.T.; Zuidema, P.A.; Sterck, F. Explaining biomass growth of tropical canopy trees: The importance of sapwood. *Oecologia* **2015**, *177*, 1145–1155. [CrossRef] [PubMed]

2. Sumida, A.; Watanabe, T.; Miyaura, T. Interannual variability of leaf area index of an evergreen conifer stand was affected by carry-over effects from recent climate conditions. *Sci. Rep.* **2018**, *8*, 13590. [CrossRef] [PubMed]

3. Kim, J.; Ryu, Y.; Jiang, C.; Hwang, Y. Continuous observation of vegetation canopy dynamics using an integrated low-cost, near-surface remote sensing system. *Agric. For. Meteorol.* **2019**, *264*, 164–177. [CrossRef]

4. de Wit, C.T. Photosynthesis of Leaf Canopies. 1965. Available online: https://library.wur.nl/WebQuery/wurpubs/413358 (accessed on 11 September 2019).

5. Ross, J. *The Radiation Regime and Architecture of Plant Stands*; Springer Netherlands: New York, NY, USA, 1981.

6. Zheng, G.; Moskal, L.M. Retrieving leaf area index (LAI) using remote sensing: Theories, methods and sensors. *Sensors* **2009**, *9*, 2719–2745. [CrossRef] [PubMed]

7. Yan, G.; Hu, R.; Luo, J.; Weiss, M.; Jiang, H.; Mu, X.; Xie, D.; Zhang, W. Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives. *Agric. For. Meteorol.* **2019**, *265*, 390–411. [CrossRef]

8. Zhao, K.; García, M.; Liu, S.; Guo, Q.; Chen, G.; Zhang, X.; Zhou, Y.; Meng, X. Terrestrial lidar remote sensing of forests: Maximum likelihood estimates of canopy profile, leaf area index, and leaf angle distribution. *Agric. For. Meteorol.* **2015**, *209*, 100–113. [CrossRef]

9. Li, Y.; Su, Y.; Hu, T.; Xu, G.; Guo, Q. Retrieving 2-D leaf angle distributions for deciduous trees from terrestrial laser scanner data. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4945–4955. [CrossRef]

10. Zhu, X.; Skidmore, A.K.; Wang, T.; Liu, J.; Darvishzadeh, R.; Shi, Y.; Premier, J.; Heurich, M. Improving leaf area index (LAI) estimation by correcting for clumping and woody effects using terrestrial laser scanning. *Agric. For. Meteorol.* **2018**, *263*, 276–286. [CrossRef]

11. Hopkinson, C.; Lovell, J.; Chasmer, L.; Jupp, D.; Kljun, N.; van Gorsel, E. Integrating terrestrial and airborne lidar to calibrate a 3D canopy model of effective leaf area index. *Remote Sens. Environ.* **2013**, *136*, 301–314. [CrossRef]

12. Hancock, S.; Essery, R.; Reid, T.; Carle, J.; Baxter, R.; Rutter, N.; Huntley, B. Characterising forest gap fraction with terrestrial lidar and photography: An examination of relative limitations. *Agric. For. Meteorol.* **2014**, *189*, 105–114. [CrossRef]

13. Calders, K.; Armston, J.; Newnham, G.; Herold, M.; Goodwin, N. Implications of sensor configuration and topography on vertical plant profiles derived from terrestrial LiDAR. *Agric. For. Meteorol.* **2014**, *194*, 104–117. [CrossRef]

14. Jupp, D.L.; Culvenor, D.; Lovell, J.; Newnham, G.; Strahler, A.; Woodcock, C. Estimating forest LAI profiles and structural parameters using a ground-based laser called Echidna®. *Tree Physiol.* **2009**, *29*, 171–181. [CrossRef] [PubMed]

15. Hu, R.; Yan, G.; Mu, X.; Luo, J. Indirect measurement of leaf area index on the basis of path length distribution. *Remote Sens. Environ.* **2014**, *155*, 239–247. [CrossRef]

16. Hu, R.; Bournez, E.; Cheng, S.; Jiang, H.; Nerry, F.; Landes, T.; Saudreau, M.; Kastendeuch, P.; Najjar, G.; Colin, J.; et al. Estimating the leaf area of an individual tree in urban areas using terrestrial laser scanner and path length distribution model. *ISPRS J. Photogramm. Remote Sens.* **2018**, *144*, 357–368. [CrossRef]

17. Lovell, J.; Jupp, D.L.; Culvenor, D.; Coops, N. Using airborne and ground–based ranging lidar to measure canopy structure in Australian forests. *Can. J. Remote Sens.* **2003**, *29*, 607–622. [CrossRef]

18. Zhu, Z.; Liu, J. Unsupervised extrinsic parameters calibration for multi-bem LiDARs. In Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering, Los Angeles, CA, USA, 1–2 July 2013; pp. 1110–1113.

19. Olivier, P. Leddar Optical Time–of–Flight Sensing Technology: A New Approach to Detection and Ranging. 2015. Available online: https://dlwx5us9wukuhO.cloudfront.net/app/uploads/dlm_uploads/2016/02/Leddar-Optical-Time-of-Flight-Sensing-Technology-l.pdf (accessed on 11 September 2019).

20. Gangadharan, S.; Burks, T.F.; Schueller, J.K. A comparison of approaches for citrus canopy profile generation using ultrasonic and Leddar®sensors. *Comput. Electron. Agric.* **2019**, *156*, 71–83. [CrossRef]

21. Arnay, R.; Hernández–Aceituno, J.; Toledo, J.; Acosta, L. Laser and Optical Flow Fusion for a Non–Intrusive Obstacle Detection System on an Intelligent Wheelchair. *IEEE Sens. J.* **2018**, *18*, 3799–3805. [CrossRef]

22. Mimeault, Y.; Cantin, D. Lighting system with driver assistance capabilities. U.S. Patent No. 8,600,656, 3 December 2013.

23. Godejord, B. Characterization of a Commercial LIDAR Module for Use in Camera Triggering System. Master's Thesis, Norwegian University of Science and Technology NTNU, Taibei, Taiwan, 2018.

24. Thakur, R. Scanning LIDAR in Advanced Driver Assistance Systems and Beyond: Building a road map for next–generation LIDAR technology. *IEEE Consum. Electron. Mag.* **2016**, *5*, 48–54. [CrossRef]

25. Mimeault, Y. Parking management system and method using lighting system. U.S. Patent No. 8,723,689, 13 May 2014.

26. Hentschke, M.; Pignaton de Freitas, E.; Hennig, C.; Girardi da Veiga, I. Evaluation of Altitude Sensors for a Crop Spraying Drone. *Drones* **2018**, *2*, 25. [CrossRef]

27. Elaksher, A.F.; Bhandari, S.; Carreon-Limones, C.A.; Lauf, R. Potential of UAV lidar systems for geospatial mapping. In Proceedings of the Lidar Remote Sensing for Environmental Monitoring, San Diego, CA, USA, 6–10 August 2017; p. 104060L.

28. Bohren, J.; Foote, T.; Keller, J.; Kushleyev, A.; Lee, D.; Stewart, A.; Vernaza, P.; Derenick, J.; Spletzer, J.; Satterfield, B. Little ben: The ben franklin racing team's entry in the 2007 DARPA urban challenge. *J. Field Rob.* **2008**, *25*, 598–614. [CrossRef]

29. Muhammad, N.; Lacroix, S. Calibration of a rotating multi-beam lidar. In Proceedings of the IROS 2010: IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan, 18–22 October 2010; pp. 5648–5653.

30. Atanacio-Jiménez, G.; González-Barbosa, J.-J.; Hurtado-Ramos, J.B.; Ornelas-Rodríguez, F.J.; Jiménez-Hernández, H.; García–Ramirez, T.; González-Barbosa, R. LIDAR velodyne HDL–64E calibration using pattern planes. *Int. J. Adv. Rob. Syst.* **2011**, *8*, 59. [CrossRef]

31. Levinson, J.; Thrun, S. Unsupervised Calibration for Multi–beam Lasers. In Proceedings of the Experimental Robotics: The 12th International Symposium on Experimental Robotics, Delhi, India, 18–21 December 2010; p. 179.

32. Sheehan, M.; Harrison, A.; Newman, P. Self-calibration for a 3D laser. *Int. J. Rob. Res.* **2012**, *31*, 675–687. [CrossRef]

33. Li, J.; He, X.; Li, J. 2D LiDAR and camera fusion in 3D modeling of indoor environment. In Proceedings of the 2015 National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, 15–19 June 2015; pp. 379–383.

34. Budge, S.E.; Badamikar, N.S.; Xie, X. Automatic registration of fused lidar/digital imagery (texel images) for three–dimensional image creation. *Optical Engineering* **2014**, *54*, 031105. [CrossRef]

35. Bodensteiner, C.; Hübner, W.; Jüngling, K.; Solbrig, P.; Arens, M. Monocular camera trajectory optimization using LiDAR data. In Proceedings of the Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 7 November 2011; pp. 2018–2025.

36. Park, Y.; Yun, S.; Won, C.S.; Cho, K.; Um, K.; Sim, S. Calibration between color camera and 3D LIDAR instruments with a polygonal planar board. *Sensors* **2014**, *14*, 5333–5353. [CrossRef] [PubMed]

37. Zhou, L.; Deng, Z. Extrinsic calibration of a camera and a lidar based on decoupling the rotation from the translation. In Proceedings of the Intelligent Vehicles Symposium (IV), Alcalá de Henares, Spain, 3–7 June 2012; pp. 642–648.

38. Fremont, V.; Bonnifait, P. Extrinsic calibration between a multi–layer lidar and a camera. In Proceedings of the 2008 IEEE International Conference on MFI, Heidelberg, Germany, 3–6 September 2006; pp. 214–219.

39. Debattisti, S.; Mazzei, L.; Panciroli, M. Automated extrinsic laser and camera inter-calibration using triangular targets. In Proceedings of the Intelligent Vehicles Symposium (IV), Gold Coast, Australia, 23–26 June 2013; pp. 696–701.

40. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, England, 2003.

41. De Silva, V.; Roche, J.; Kondoz, A. Robust fusion of LiDAR and wide-anglse camera data for autonomous mobile robots. *Sensors* **2018**, *18*, 2730. [CrossRef] [PubMed]

42. Jia, Q.; Fan, X.; Luo, Z.; Song, L.; Qiu, T. A fast ellipse detector using projective invariant pruning. *IEEE Trans. Image Process.* **2017**, *26*, 3665–3679. [CrossRef] [PubMed]

43. Fitzgibbon, A.W.; Pilu, M.; Fisher, R.B. Direct least squares fitting of ellipses. In Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria, 25–29 Aug 1996; pp. 253–257.

44. Newton, I. De analysi per aequationes numero terminorum infinitas. 1711.

45. Newton, I.; Colson, J. *The Method of Fluxions and Infinite Series; with Its Application to the Geometry of Curve-lines... Translated from the Author's Latin Original Not Yet Made Publick. To which is Subjoin'd a Perpetual Comment Upon the Whole Work... by J. Colson*; Henry Woodfall: London, UK, 1736.

46. Ypma, T.J. Historical development of the Newton-Raphson method. *SIAM Rev.* **1995**, *37*, 531–551. [CrossRef]

47. Huber, P.J. *Robust Statistics*; Wiley Online Library: New York, NY, USA, 1981; p. ix. 308p. Available online: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470434697 (accessed on 11 September 2019).

48. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. In Proceedings of the European conference on computer vision, Graz, Austria, 7–13 May 2006; pp. 404–417.

49. Nock, R.; Nielsen, F. Statistical region merging. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1452–1458. [CrossRef]

50. Pertuz, S.; Kamarainen, J. Region-based depth recovery for highly sparse depth maps. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 2074–2078.

51. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [CrossRef]

52. Zanewich, K.P.; Pearce, D.W.; Rood, S.B. Heterosis in poplar involves phenotypic stability: Cottonwood hybrids outperform their parental species at suboptimal temperatures. *Tree Physiol.* **2018**, *38*, 789–800. [CrossRef]

53. Chen, Y.; Zhang, W.; Hu, R.; Qi, J.; Shao, J.; Li, D.; Wan, P.; Qiao, C.; Shen, A.; Yan, G. Estimation of forest leaf area index using terrestrial laser scanning data and path length distribution model in open–canopy forests. *Agric. For. Meteorol.* **2018**, *263*, 323–333. [CrossRef]

54. Schleppi, P.; Conedera, M.; Sedivy, I.; Thimonier, A. Correcting non-linearity and slope effects in the estimation of the leaf area index of forests from hemispherical photographs. *Agric. For. Meteorol* **2007**, *144*, 236–242. [CrossRef]

55. Thimonier, A.; Sedivy, I.; Schleppi, P. Estimating leaf area index in different types of mature forest stands in Switzerland: A comparison of methods. *Eur. J. For. Res.* **2010**, *129*, 543–562. [CrossRef]

56. Lang, A. Simplified estimate of leaf area index from transmittance of the sun's beam. *Agric. For. Meteorol.* **1987**, *41*, 179–186. [CrossRef]

57. Lang, A.; Xiang, Y. Estimation of leaf area index from transmission of direct sunlight in discontinuous canopies. *Agric. For. Meteorol.* **1986**, *37*, 229–243. [CrossRef]

58. Strahler, A.H.; Jupp, D.L.B.; Woodcock, C.E.; Schaaf, C.B.; Yao, T.; Zhao, F.; Yang, X.; Lovell, J.; Culvenor, D.; Newnham, G.; et al. Retrieval of forest structural parameters using a ground–based lidar instrument (Echidna®). *Can. J. Remote Sens.* **2008**, *34*, S426–S440. [CrossRef]

59. Pimont, F.; Soma, M.; Dupuy, J.-L. Accounting for Wood, Foliage Properties, and Laser Effective Footprint in Estimations of Leaf Area Density from Multiview–LiDAR Data. *Remote Sens.* **2019**, *11*, 1580. [CrossRef]