

Article

Extracting Features from Oblique Ground-Based Multispectral Images for Monitoring Cotton Physiological Response to Nitrogen Treatments

Vaishali Swaminathan ^{1,2,*}, J. Alex Thomasson ^{1,3} , Nithya Rajan ⁴ and Robert G. Hardin ¹

- ¹ Department of Biological and Agricultural Engineering, Texas A&M University, College Station, TX 77843, USA; athomasson@abe.msstate.edu (J.A.T.); robert.hardin@ag.tamu.edu (R.G.H.)
² Utah Water Research Laboratory, Utah State University, Logan, UT 84321, USA
³ Department of Agricultural and Biological Engineering, Mississippi State University, Starkville, MS 39762, USA
⁴ Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843, USA; nithya.rajan@ag.tamu.edu
* Correspondence: vaishaliswaminathan@tamu.edu

Abstract: Early detection of nitrogen deficiency in cotton requires timely identification of stress symptoms like leaf chlorosis (yellowing) and canopy stunting. Chlorosis initially appears in older, lower-canopy leaves, which are often not visible in conventional nadir-looking imaging. This study investigates oblique ground-based multispectral imaging to estimate plant height and capture spectral details from the upper (UC) and lower (LC) cotton canopy layers. Images were collected from four camera pitch and height configurations: set 1 (30°, 2 m), set 2 (55°, 2 m), set 3 (68°, 3 m), and set 4 (70°, 1.5 m). A pre-trained monocular depth estimation model (MiDaS) was used to estimate plant height from aligned RGB images and an empirically derived tangential model corrected for perspective distortion. Further, the lower and upper vertical halves of the plants were categorized as LC and UC, with vegetation indices (CI_{green} , $CI_{rededge}$) calculated for each. The aligned images in set 1 had the best sharpness and quality. The plant height estimates from set 1 had the highest correlation ($r = 0.64$) and lowest root mean squared error (RMSE = 0.13 m). As the images became more oblique, alignment and monocular depth/height accuracy decreased. Also, the effects of perspective and object-scale ambiguity in monocular depth estimation were prominent in the high oblique and relatively low altitude images. The spectral vegetation indices (VIs) were affected by band misalignment and shadows. VIs from the different canopy layers demonstrated moderate correlation with leaf nitrogen concentration, and sets 2 and 3 specifically showed high and low differences in VIs from the UC and LC layers for the no and high-nitrogen treatments, respectively. However, improvements in the multispectral alignment process, extensive data collection, and ground-truthing are needed to conclude whether the LC spectra are useful for early nitrogen stress detection in field cotton.

Keywords: terrestrial platforms; remote sensing; computer vision; height from depth



Academic Editor: Kevin Tansey

Received: 9 December 2024

Revised: 17 January 2025

Accepted: 5 February 2025

Published: 8 February 2025

Citation: Swaminathan, V.; Thomasson, J.A.; Rajan, N.; Hardin, R.G. Extracting Features from Oblique Ground-Based Multispectral Images for Monitoring Cotton Physiological Response to Nitrogen Treatments.

Remote Sens. **2025**, *17*, 579. <https://doi.org/10.3390/rs17040579>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A wide range of ground-based platforms have been used in agriculture for data collection, fertilizer application, weed removal, harvesting, etc., and they include machines like push carts, motorized vehicles, and unmanned or robotic ground platforms. Sensing from ground-based platforms generally falls under the category of proximal sensing, where

the sensors are directly in contact with or within a few meters of the plants [1]. These platforms have a higher payload capacity than the commonly used unmanned aerial vehicles (UAVs), and compared to remote sensing, proximal sensing is less susceptible to atmospheric noise and provides features with higher resolutions that enable organ-level analyses of plants. While ground-based platforms can be adapted with relative ease to implement digital sensing technologies, data collection with such platforms can be time-consuming.

Ground-based platforms are commonly equipped with various sensors, cameras, GPS/GNSS, data acquisition systems, computing units, and actuators. Multi-sensor platforms that monitor various physiological and spectral plant characteristics are particularly useful for precision nitrogen (N) management, as N availability and uptake impact not only the leaf reflectance spectra but also the growth and development of canopies. In an earlier study by Sui and Thomasson [2], a neural network model based on ground-based multispectral images and ultrasonic height measurements was effective in determining cotton leaf N concentration and predicted N stress with about 90% accuracy. Multi-sensor mobile platforms are also used in phenotyping, early-season crop mapping, general biotic or abiotic stress monitoring, and yield estimation [3]. The Robotanist is an autonomous sorghum phenotyping platform equipped with sonar proximity sensors, IR radiometers, and NIR cameras for canopy height, temperature, and health measurements [4]. However, with advancements in computer vision, stereovision cameras with depth perception capabilities are viable alternatives in many cases for LiDAR or ultrasonic sensors. Further, photogrammetric depth estimation techniques and deep learning models can compute height, depth, and other 3D volumetric features from the perceived overlap between images from conventional RGB, multispectral, or hyperspectral cameras. Since photogrammetric image processing facilitates the extraction of spectral and morphological plant features from images, it minimizes the sensor load and cost. Examples of more advanced ground-based platforms include the GPhenoVision, which is a manually driven cotton phenotyping vehicle equipped with hyperspectral, thermal, and depth cameras [5]. Issac et al. [6] developed a high-throughput cotton phenotyping platform with cloud computing and object detection capabilities that use the YOLOv5 model. Francesca et al. [7] developed a low-cost, non-motorized tomato phenotyping vehicle with cameras mounted on a bicycle chassis. Such platforms are lightweight and so do not compact the soil, and they are generally used for smaller plants like tomatoes and strawberries. Ground-based platforms are also used in see-and-spray weed management [8] and robotic harvesting [9].

Feature matching is a fundamental component of classical computer vision, used in various applications like image alignment, epipolar rectification for stereo vision, and photogrammetric 3D reconstruction. Scale-invariant feature transformation (SIFT) is a commonly used local feature descriptor that is robust and independent of scale, orientation, and distortions in images [10]. Studies that compare SIFT with other feature descriptors such as sped-up robust feature (SURF), binary robust independent features (BRIEF), features from accelerated segment test (FAST), oriented FAST and rotated BRIEF (ORB), etc., have shown that SIFT had the best accuracy under most conditions, even though the other algorithms were faster [11,12]. Simultaneous localization and mapping (SLAM) algorithms that combine image features with GPS and IMU data were used to navigate through and monitor cotton fields [13]. The AGRI-SLAM algorithm [14] used ORB features to compute epipolar transformation between stereo image pairs, whereby this SLAM implementation based on only visual inputs can operate in regions with poor GPS networks. Another study used depth thresholding from disparity images to classify the ridges and furrows in fields and eventually determine guidance lines for autonomous navigation [15]. Stereovision systems and image segmentation have been useful for phenotyping applications like

determining the canopy height of various plant species [16]. Instead of commercial stereo cameras, Bao et al. [17] used RGB cameras to build custom stereovision systems that estimated 3D sorghum plant features such as plant height, stem diameter, leaf area, leaf count, panicle size, etc.

Aside from classical computer vision, deep learning models have become popular for automatic object detection, feature extraction, and image segmentation. Deep learning models like RegiNet [18] and neural style transfer (NST) [19] have been trained for multispectral image alignment. Deep learning models with a convolutional neural network (CNN) framework have been used in general agricultural object detection and tracking [20], precision weed management [21] and robotic harvesting [22] from ground-based platforms. Object detection with CNN models on ground-based images has been used for mango yield estimation and mapping [23,24]. Sethy et al. [25] passed feature vectors from the CNN layers to a support vector machine (SVM) regression model to estimate nitrogen (N) deficiency levels from UAV images. Monocular depth estimation, as an alternative to stereovision, uses deep-learning CNN models to estimate pixel-level depth from single RGB images. Zhao et al. [26] developed a ResNet model for monocular depth estimation, which was subsequently used to estimate plant height and width. Many of these studies highlighted the need for appropriate image distortion correction to reliably reconstruct 3D canopy models. While some monocular depth estimation models such as monoDA [27] and P3ES-Net [26] were specifically designed for agricultural datasets, general purpose depth estimation models like MiDaS (monocular depth estimation via a multi-scale network) [28,29] and MonoViT [30] can be used in agriculture, too. The MiDaS model, used in this study, has a dense prediction transformer (DPT) architecture that uses encoder–decoder networks to hierarchically extract multi-scale features and capture finer pixel-level details for segmentation and depth estimation. The model was trained with diverse datasets and demonstrated capabilities for zero-shot transfer learning, i.e., to predict datasets it was not familiar with during training. While the studies by Tangellapalli et al. (2024) [31] for monocular depth estimation in rainforests and Mizuta et al. (2024) [32] for canopy coverage estimation in wheat used MiDaS to model images acquired from UAV platforms, its performance on oblique ground-based images remains unexplored.

Most of the studies cited above followed the conventional approach of capturing images with nadir-looking cameras and occasionally used forward-looking cameras in special cases like robotic harvesting, where the objects were in front of the camera. There is also a growing interest in the use of oblique images for capturing details from the sides and bottom of the object that are generally obscured in nadir images [33]. One phenotyping study achieved organ-level reconstruction of plant canopies with UAV-based multi-view oblique photography [34]. A similar approach was followed for organ-level reconstructions from very low altitude (3 m above ground) oblique images, in which the 3D model could clearly identify the leaves and fruiting organs of maize, soybean, and tobacco plants in mixed cropping systems [35]. The leaves were also categorized as upper, middle, or lower based on their vertical positioning along the stem. Che et al. [36] showed that oblique UAV images could be used to reconstruct the entire canopy structure better than nadir images; however, photogrammetric reconstruction and estimated plant height accuracy were best when the nadir and oblique images were combined. Li et al. [37] was one of the few studies that implemented oblique imaging from a ground-based platform, in which oblique imaging showed promising results for wheat canopy reconstruction from a single camera. Other ground-based oblique imaging applications have included wheat spike detection [38] and cotton boll detection [6]. In the context of precision N management, the lower canopy spectral response is crucial because leaf senescence starts with the older leaves before progressing to newer leaves. When subjected to N stress, plants remobilize

nutrients from the older leaves in the lower canopy to the newer ones at the top to sustain photosynthesis. Since nadir-looking images only capture top-of-canopy reflectance, parts of the canopy that may show early symptoms of N stress are overlooked.

In this study, we proposed using oblique ground-based multispectral imaging for the extraction of plant spectral and physical features that may be early indicators of N stress in cotton. The goal was to determine whether oblique images can capture physiological changes, specifically the rapid remobilization of N from the lower to upper canopy, in response to N deficiency. The specific objectives to accomplish the goal were (1) to develop an alignment workflow for the different bands of a multispectral camera, (2) to estimate plant height through monocular depth estimation, and (3) to analyze the spectral features in the upper and lower canopy layers through vertical canopy segmentation.

2. Materials and Methods

A computer vision workflow was conceptualized for extracting cotton-canopy related spectral features and plant height from ground-based oblique images. Oblique viewing angles were specifically chosen to determine whether the bottom layers of the canopy provided useful information pertaining to early symptoms of N stress. This study used a simple image-acquisition setup that can be achieved with any existing multispectral camera and ground-based agricultural platform. Some of the challenges in deriving quantitative features from oblique images, like handling perspective-induced distortion during image alignment and canopy height estimation, were also addressed. The overall workflow consisted of band alignment, undistorted canopy height extraction, vertical canopy segmentation, and an analysis of spectral vegetation indices (VIs) in different canopy layers (Figure 1). The two output features were undistorted above-ground canopy height and VIs at different canopy layers.

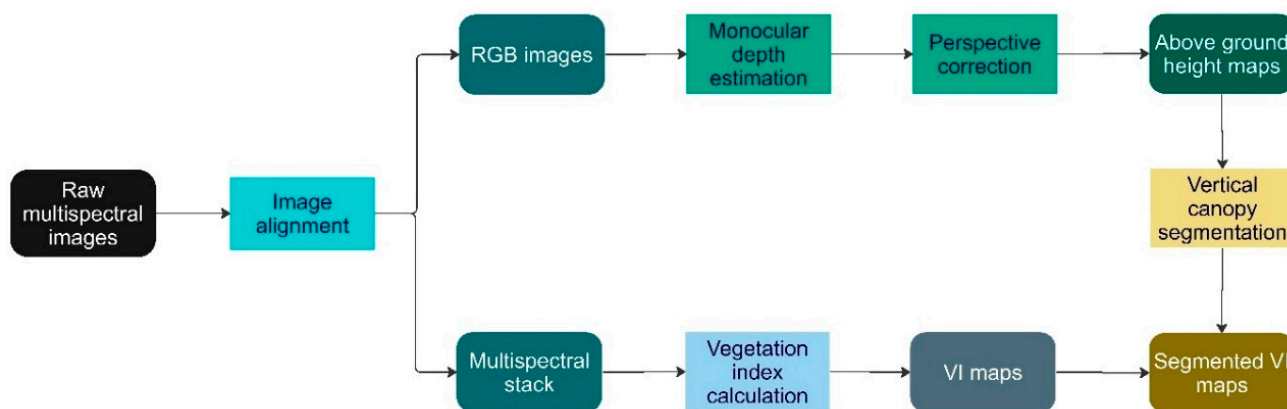


Figure 1. General workflow of steps involved in processing ground-based oblique multispectral images.

2.1. Experimental Setup

This experiment was conducted in 2022 in a cotton field located at the Texas A&M Agrilife Research farm near College Station, TX, USA. A cotton N management trial was conducted, which consisted of three replications with six treatments (18 plots in total) set up as a randomized block design. In each treatment plot, the center four rows were used for the N management study and the two bordering rows on either side were considered as buffers between adjacent plots (Figure 2). The six N treatments were administered in two split applications as follows: 0 + 0 kg/ha (T1; no N), 0 + 56 kg/ha (T2; low N), 0 + 112 kg/ha (T3; optimal N), 56 + 0 kg/ha (T4; low N), 112 + 0 kg/ha (T5; optimal N), and 56 + 112 kg/ha (T6; high N). The initial N application was performed 30 days after planting (DAP), and the second application was made at 44 DAP, around the time of pin-head square

formation. Image data were collected at 50 DAP, and the ground truth for the plant height and biomass analyses were collected at 52 DAP. The leaves from the biomass samples were dried until a constant dry weight was achieved, after which they were grounded and analyzed for leaf nitrogen concentration (N %) through combustion [39].

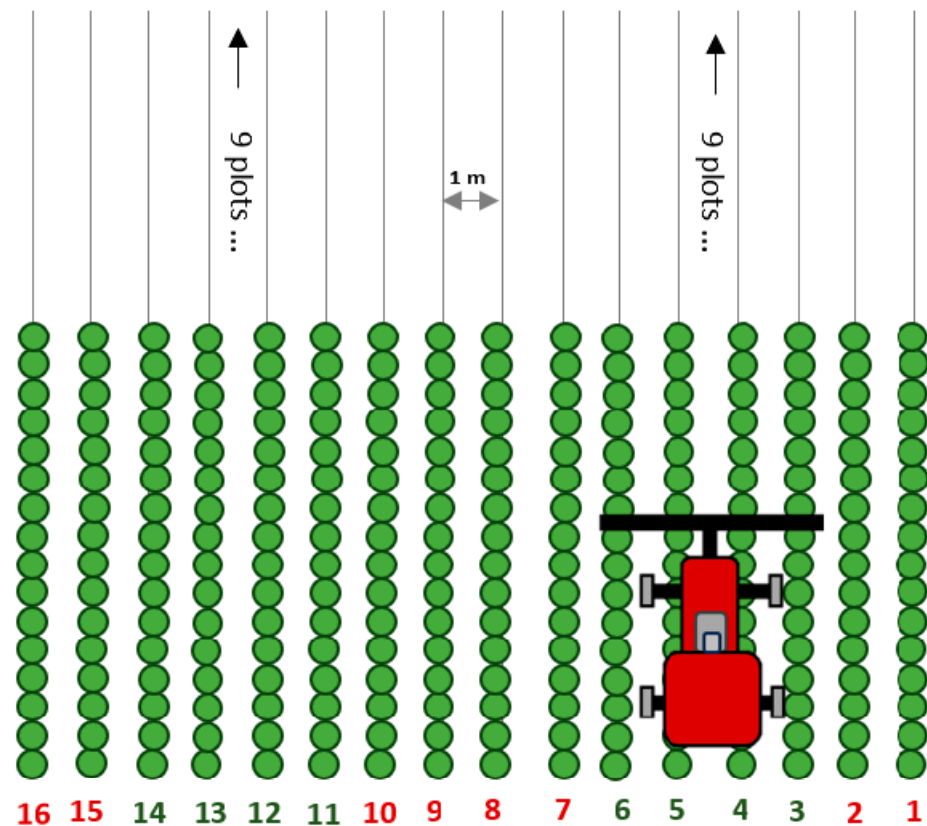


Figure 2. A representation of the data collection process of the ground-based cotton phenotyping platform. The platform covered the center four rows (numbers in green) of the plots during each pass, while skipping the two border rows (numbers in red) on each side of the plot.

2.2. Platform Description

A custom-built motorized cotton phenotyping platform was used for data collection. The wheels of the platform were roughly 2 m apart, and the center two rows of the plots fit between the wheels (Figure 2). The platform's front sensor panel covered four rows at a time. Since the platform covered only nine plots in one pass, it made two passes to cover all eighteen plots in the experiment (Figure 2). A Micasense RedEdge-MX (AgEagle Aerial Systems, Wichita, KS, USA) five-band multispectral camera was mounted with a cellphone holder onto an extended arm attached to a vertical beam on the platform (Figure 3b). The arm was moved up and down the vertical beam to adjust the camera's above-ground mounting height. Although there was limited freedom for angular adjustment of the arm, the cellphone holder attached to the end of the arm provided additional rotational movement for adjusting the camera's pitch. Micasense's downwelling light sensor (DLS) with built-in GPS and IMU modules was attached to the back of the camera to record the camera's pitch at the time of image acquisition. Four sets of oblique images of the crop rows were acquired with auto-exposure settings at 1.0 s intervals with different combinations of camera pitch and above-ground mounting height, each labeled as a particular 'set' (Table 1). The camera manufacturer provided a calibrated reflectance panel (CRP) (Figure 3c) for image reflectance calibration, and an elevated GCP was used for height calibration (Figure 3a). Five images representing each plot were randomly sampled from the entire dataset for analysis.

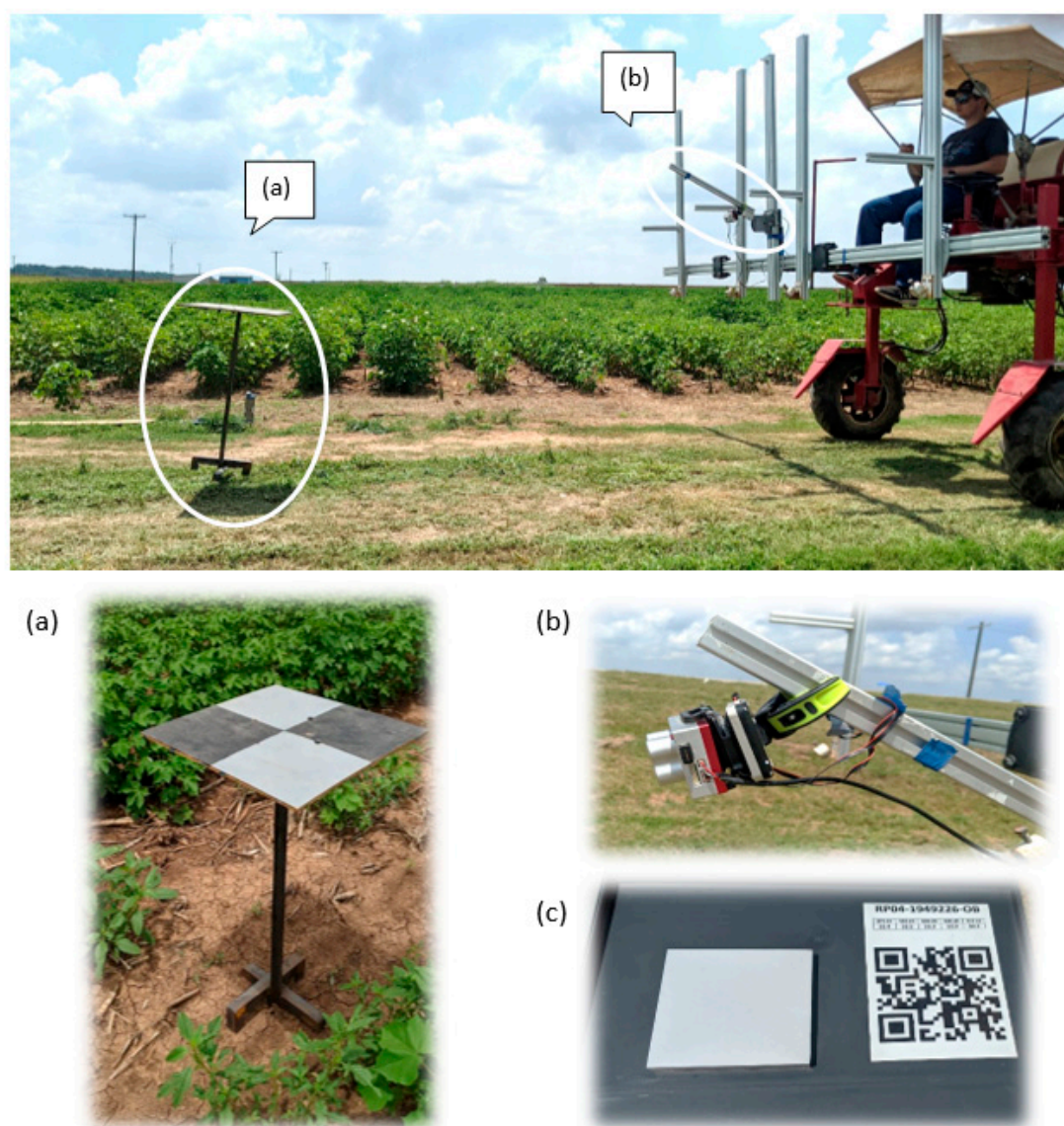


Figure 3. Instrumentation and calibration setup for ground-based image collection. (a) Elevated GCPs for height calibration, (b) Micasense RedEdge-MX camera mounted on an adjustable arm attached to the vertical beam on the vehicle, and (c) Micasense calibrated reflectance panel (CRP) for reflectance calibration.

Table 1. Camera mounting height above-ground (m) and camera pitch ($^{\circ}$) combinations for different image datasets.

Dataset Label	Above-Ground Camera Mounting Height (m)	Camera Pitch ($^{\circ}$)
Set 1	2.0	30
Set 2	2.0	55
Set 3	3.0	68
Set 4	1.5	70

2.3. Band Alignment

The five bands of the multispectral images were not stacked directly on top of one another due to differences in lens orientation. Firstly, the images were corrected for lens distortion by replicating the algorithm available in the Micasense image processing library (<https://github.com/micasense/imageprocessing> (accessed on August 2022)). The algorithm removed barrel distortion in the raw images with the help of the camera matrix

and distortion parameters extracted from the image metadata. The camera matrix was defined by the focal length (f) and principal point coordinates (c_x, c_y) (Equation (1)), and the lens distortion vector was described by three pre-calibrated radial (k_1, k_2 , and k_3) and two tangential (p_1 and p_2) distortion coefficients (Equation (2)). The red-edge band was used as references for warping and aligning the five bands because the red-edge lens was housed in the center of the camera and was equidistant to all the other lenses. In Micasense's original algorithm for image alignment, the warping matrix was determined by the enhanced correlation coefficient (EEC) image alignment method [40] that iteratively optimized the matrix using Gaussian blurring, also known as Gaussian pyramid features. This method is effective in terms of alignment accuracy and computation time for nadir-looking UAV images captured at higher altitudes, as the warping matrix computed for one image can be applied to the entire dataset as long as the camera orientation remains unchanged. However, applying a uniform matrix for ground-based images resulted in misalignment of the bands due to variations in camera orientations induced by the platform's vibrational motion while traversing the field. Also, computing the EEC matrix for each image in the dataset is computationally intensive and slow. Hence, this study adopted the k-nearest-neighbor (kNN) SIFT-descriptor matching followed by homography transformation to determine the warping matrix for each instance. For each descriptor in the reference image, the kNN algorithm with $k = 2$ matched two descriptors in the target image with the least Euclidean distance to the reference. To ensure the matches were unique, only the descriptor pairs with Euclidean distance less than 0.7 times the distance of the next best pair were selected. Random sampling of consensus (RANSAC) was used iteratively to eliminate outliers during homography transformation. Perspective transformation was performed with the homography matrices computed for each band in the image to create aligned RGB and five-band multispectral stacks. The reflectance calibration coefficients derived from the CRP images were used to calibrate the multispectral reflectance stacks. Since no high-quality reference images were available for image-quality comparisons after alignment, the blind/referenceless image spatial quality evaluator, also known as the BRISQUE score, was used to determine the alignment quality. BRISQUE is a pre-trained SVR model that returns a quality assessment score based on distortions in images, especially when a high-quality undistorted reference image is not available [41]. Also, a high-pass (Laplacian) filter was applied to the aligned RGB images, where higher variance indicated a low level of blur or high degree of sharpness.

$$\text{camera matrix} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$\text{distortion parameters} = [k_1, k_2, p_1, p_3, k_3] \quad (2)$$

2.4. Plant Height Estimation

The monocular depth estimation model (MiDaS) with pre-trained weights was used to estimate inverse depth (D_{est}) from the oblique RGB images. The pre-trained weights (dpt_swin2_large_384) [28], which were optimized for better speed, resulted in moderate quality depth estimates. The depth estimation process was boosted with an NVIDIA RTX-3080 GPU. To correct perspective distortion in the inverse depth maps caused by the oblique viewing angles, an inverse function was empirically estimated and applied to scale each image row based on its pitch relative to the camera's optical center (Figure 4). The relative camera pitch for each row in the image (θ_i) was derived from the known camera pitch (θ_p), vertical field of view (VFOV, θ_{VFOV}), and distance of the i^{th} row (d_i) from the center

(C) (Equation (3)). A scaled depth map (D_{sc}), which was corrected for perspective, was obtained by scaling D_{est} with the tangent of θ_i (Equation (4)).

$$\theta_i = \theta_P + i * \theta_{VFOV} / n \quad (3)$$

$$D_{sc} = \tan\theta_i * D_{est} \quad (4)$$

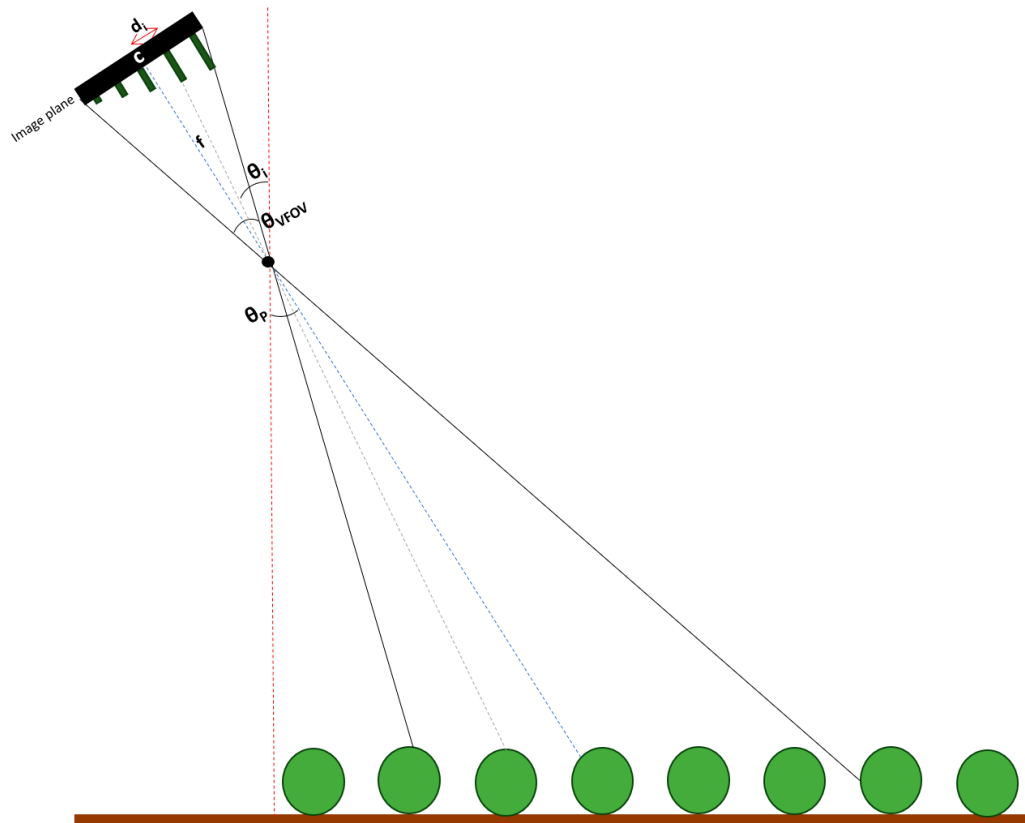


Figure 4. Perspective distortion in oblique images where objects closer to the camera appear larger and objects farther away appear smaller.

The RGB images were transformed into the hue–saturation–value (HSV) color space, where the reddish-brown soil pixels were identified as those having HSV values between [10, 0, 0] and [20, 255, 255]. With the soil mask, 100 points representing the ground surface were randomly sampled from D_{sc} to linearly interpolate the ground surface. Then, an above-ground canopy height map (h_{sc}) was obtained by subtracting the ground surface map from D_{sc} . The average above-ground plant height for each plot was determined by sampling five points representing the tallest part of the plant from five images per plot. Since h_{sc} and the sampled above-ground height values were unitless entities, a scaling factor was derived from the height calibration target to convert the unitless estimated height to metric unit (m). The accuracy of the image-estimated average plant height was evaluated by comparing it to the ground-truth values using Pearson’s correlation coefficient (r) and root mean square error (RMSE) metrics.

2.5. Vertical Plant Canopy Segmentation and Spectral Analysis

From h_{sc} , pixels in the top and bottom halves of the plant canopy were segmented and classified as the upper canopy (UC) or lower canopy (LC) layers. The ground surface was masked out to avoid misclassification of soil pixels as LC. For localized canopy segmentation, a 20-pixel tall sliding window was used to scan through each row of the height map. The horizontal width of the window was manually assigned according to the number

of crop rows captured in the images. For example, images with four or fewer crop rows were assigned 600×20 pixel windows, and images with more than four crop rows were assigned 200×20 windows. This ensured that the sliding window did not occupy more than one row at a time, which could result in incorrect segmentation of the UC and LC layers. The maximum estimated canopy height (h_{max}) was determined within each window, and the pixels were classified according to Equation (5).

$$canopy\ segmentation(h) = \begin{cases} LC, h < 0.5 * h_{max} \\ UC, h \geq 0.5 * h_{max} \end{cases} \quad (5)$$

Two chlorophyll-sensitive VIs, CI_{green} (Equation (6)) and $CI_{rededge}$ (Equation (7)), were calculated from the aligned multispectral reflectance stacks, and their average values in the LC and UC layers were determined.

$$CI_{green} = \frac{\rho_{NIR}}{\rho_{green}} - 1 \quad (6)$$

$$CI_{rededge} = \frac{\rho_{NIR}}{\rho_{rededge}} - 1 \quad (7)$$

3. Results

3.1. Perspective Correction and Plant Height Estimation

The monocular depth estimated from oblique images were prone to perspective-induced distortion as the relative distance of each object to the camera varied within the scene. An exponential distortion in inverse depth (D_{est}) was observed, whereby the depth decreased from the front to the back of the depth maps for all height–pitch combinations (Figure 5). Also, the differences in depth between the soil and canopy pixels diminished progressively due to perspective, which affected the above-ground canopy height estimation. However, the empirically derived perspective scaling factor ($\tan \theta_i$) effectively compensated for the exponential decrease in depth and diminishing above-ground canopy height, which resulted in uniform estimation of plant height (h_{sc}) from the front to back of the images (Figure 6). Although, in cases with extreme perspective, due to high camera pitch and/or low camera mounting height (set 4), the empirical correction factor was only partially effective, resulting in diminishing canopy height values towards the back (Figure 6d). The monocular depth estimation and above-ground plant height estimation processes were also prone to random errors. For example, Figure 7 shows errors due to object-scale ambiguity, where the estimated heights of the calibration target differed based on its relative distance from the camera, and the differences increased as the pitch increased from sets 1 to 4. Also, in a few cases, the ground-surface interpolation method for above-ground height estimation produced uneven ground surface maps that resulted in scaled height (h_{sc}) maps with visible artifacts and above-ground plant height estimation errors (Figure 8). Due to the errors in depth estimation and surface interpolation, the image-estimated plant height showed moderate ($r = 0.81$) to weak ($r = 0.07$) correlation with the ground-truth estimates of plant height. Figure 9 compares the plot averages of the estimated above-ground plant height with the ground-truth plant height measurements. The actual and estimated plant height measurements were analyzed separately for each pass (passes 1 and 2) and for both passes combined (pass 1 + 2). Due to object-scale ambiguity, better correlation and lower RMSE were observed when the two passes were analyzed separately. The overall correlation for both passes combined was moderate for sets 1 and 4 ($r = 0.47$ and 0.57); however, the RMSE values increased progressively from sets 1 to 4, possibly due to the

increasing uncertainty in estimating the height of the calibration targets and the occasional ground surface reconstruction errors.

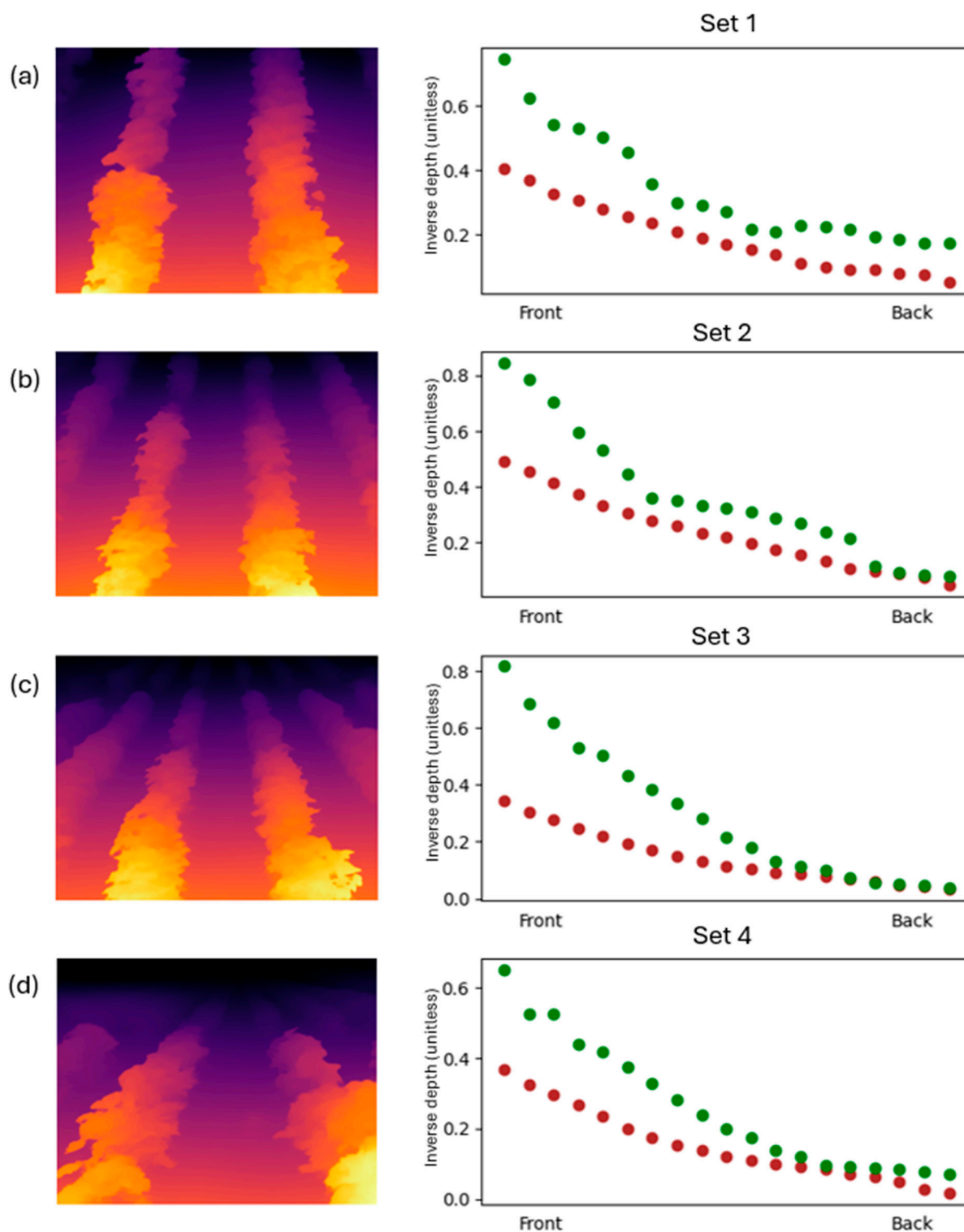


Figure 5. (Left) Inverse depth maps generated from RGB images through monocular depth estimation and (right) the corresponding inverse depth (unitless) values of canopy (green dots) and soil (brown dots) observed to decrease from the front to the back of the images for (a) set 1 (2 m—30°), (b) set 2 (2 m—55°), (c) set 3 (3 m—68°), and (d) set 4 (1.5 m—70°).

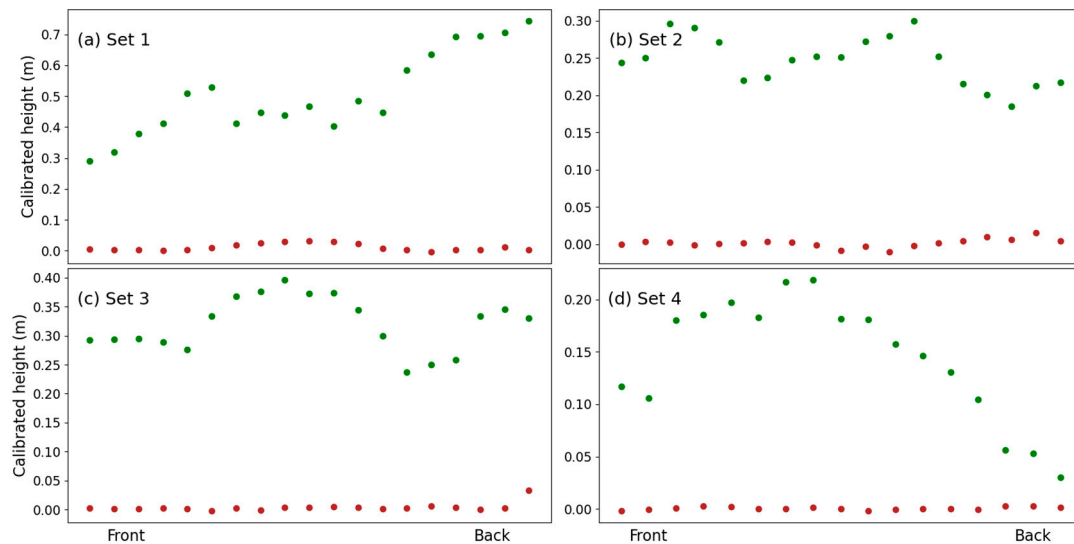


Figure 6. Empirically scaled above-ground height (h_{sc} , m) of canopy (green dots) and soil (brown dots) pixels from the front to the back of the images in (a) set 1 (2 m—30°) (top left), (b) set 2 (2 m—55°) (top right), (c) set 3 (3 m—68°) (bottom left), and (d) set 4 (1.5 m—70°) (bottom right).

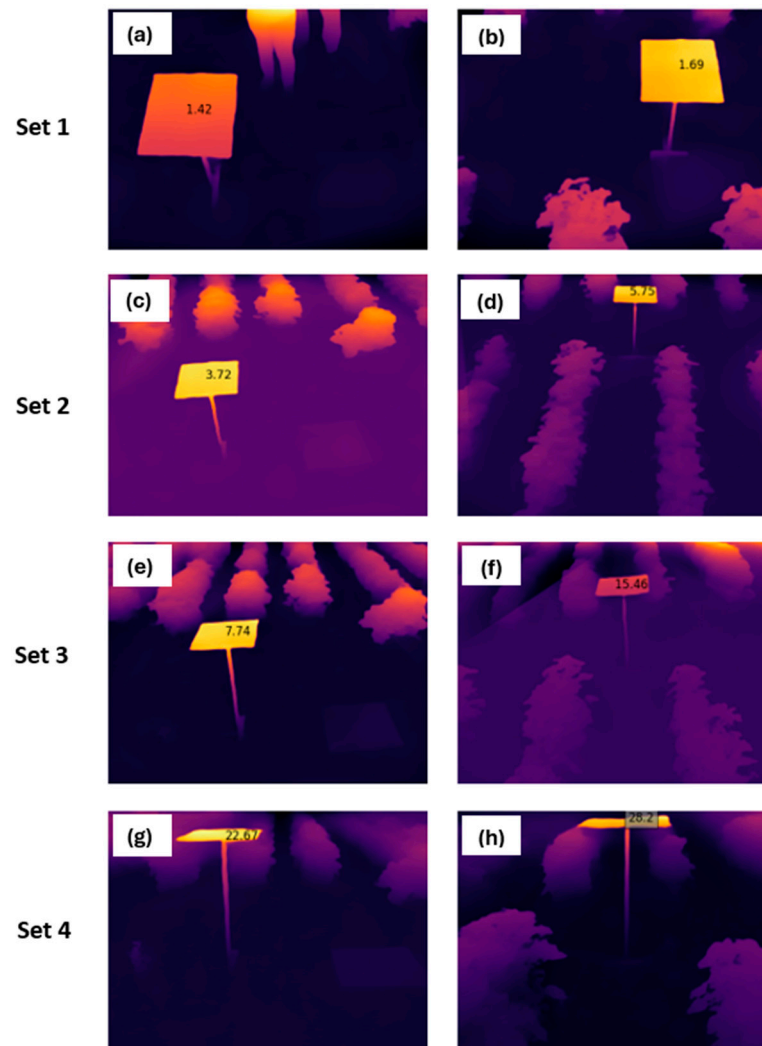


Figure 7. The above-ground height values of a calibration target of known height placed at different positions with respect to the camera for (a,b) set 1 (2 m—30°), (c,d) set 2 (2 m—55°), (e,f) set 3 (3 m—68°), and (g,h) set 4 (1.5 m—70°).

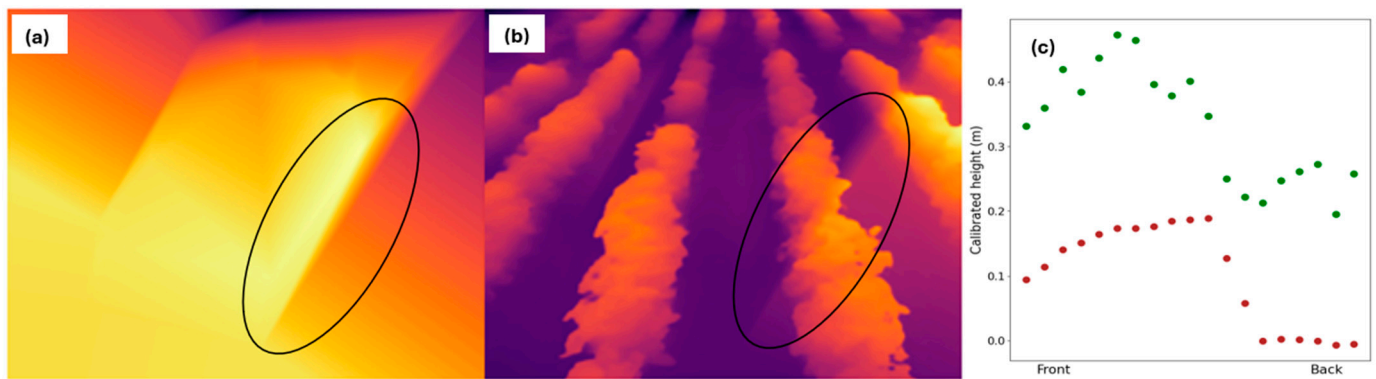


Figure 8. (a) Errors in ground surface interpolation resulted in an uneven elevation map (circled), (b) which resulted in above-ground height maps with artifacts (circled), and (c) non-uniform above-ground plant height (m) estimation (green dots) and ground surface with non-zero elevation (brown dots).

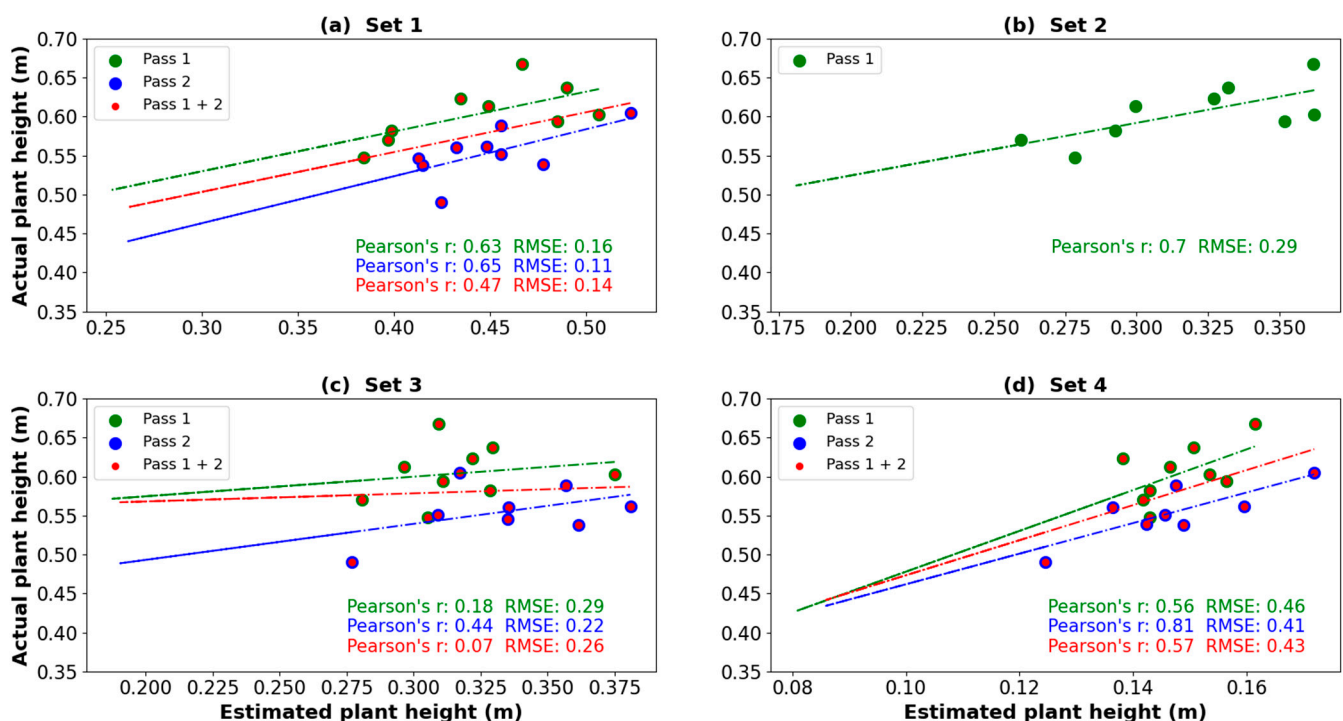


Figure 9. Comparison of image-estimated above-ground height (m) with ground-truth plant height measurements for (a) set 1 (2 m—30°) (top left), (b) set 2 (2 m—55°) (top right), (c) set 3 (3 m—68°) (bottom left), and (d) set 4 (1.5 m—70°). The green and blue dots and lines represent results from pass 1 and pass 2, respectively. The red dots and line represent results from both passes combined.

3.2. Image Alignment Quality Analysis

The five multispectral bands were stacked by performing image transformation on the blue, green, red, and NIR bands with respect to the red-edge band image. Figure 10 shows that the SIFT descriptors for all five bands mostly belonged to pixels in the canopy edges or objects on the ground, indicating that reliable features for matching could not be extracted from within the canopy. The potential error in SIFT descriptor matching was determined and estimated with the mean Euclidean distance between the SIFT descriptor pairs for the individual image bands in each run (Table 2). The feature matching was the least for all bands in set 1 and the most for bands in set 4. Set 2 with the same mounting height as set 1 but with relatively greater pitch was prone to more errors. Set 3 due to higher mounting height produced fewer errors than set 2. Set 4 due to both higher pitch and

lower mounting height resulted in the most errors. This translated into misalignment of the bands, where images in sets 1 appeared better aligned than the images in the other sets (Figure 11). With the red-edge band as a reference, the red band descriptors in all sets had the minimum error in terms of Euclidean distance. Also, the misalignment in the stacked RGB and multispectral image bands were also more pronounced in the canopy pixels than the soil pixels (Figure 11). The image quality assessment metrics corroborated the feature matching errors (Table 2) and the visual observations of misalignment (Figure 11). Images in set 1 had a low BRISQUE score (17.69), indicating a low level of misalignment, compared to images in the other sets that had relatively high BRISQUE scores (≈ 25). Since the camera had a fixed focal length, the set 3 images acquired from 3 m above-ground height had the highest sharpness score of 2064.27, and the sharpness decreased for lower camera mounting height and higher pitch. With a consistent camera height of 2 m, the image from set 1 with lower pitch (30°) was sharper than the image from set 2 with 55° pitch. In summary, image quality metrics and visual observations progressively deteriorated from set 1 to set 4 with the increase in camera pitch and decrease in mounting height.

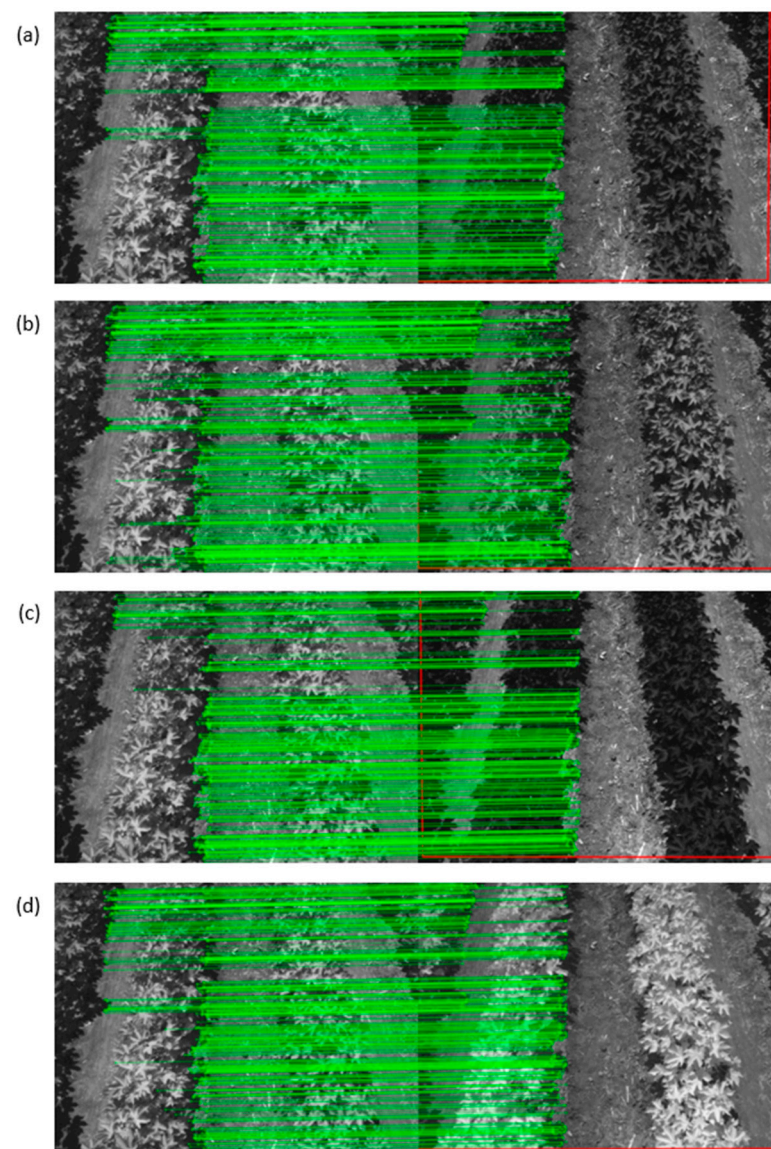


Figure 10. Scale-invariant feature transformation (SIFT) descriptor matching between the reference red edge band image (left) and (a) blue, (b) green, (c) red, and (d) NIR band images showing descriptors primarily belonging to the canopy edge or ground surface.

Table 2. SIFT feature matching errors measured in terms of mean distance between matched key points for the multispectral bands with respect to the red-edge band for set 1 (2 m—30°), (c, d) set 2 (2 m—55°), (e, f) set 3 (3 m—68°), and (g, h) set 4 (1.5 m—70°).

Bands	Set 1	Set 2	Set 3	Set 4
Blue	133.35	159.45	141.72	161.59
Green	130.01	140.20	136.46	140.22
Red	107.28	118.50	111.08	117.82
NIR	136.72	137.77	142.23	143.62

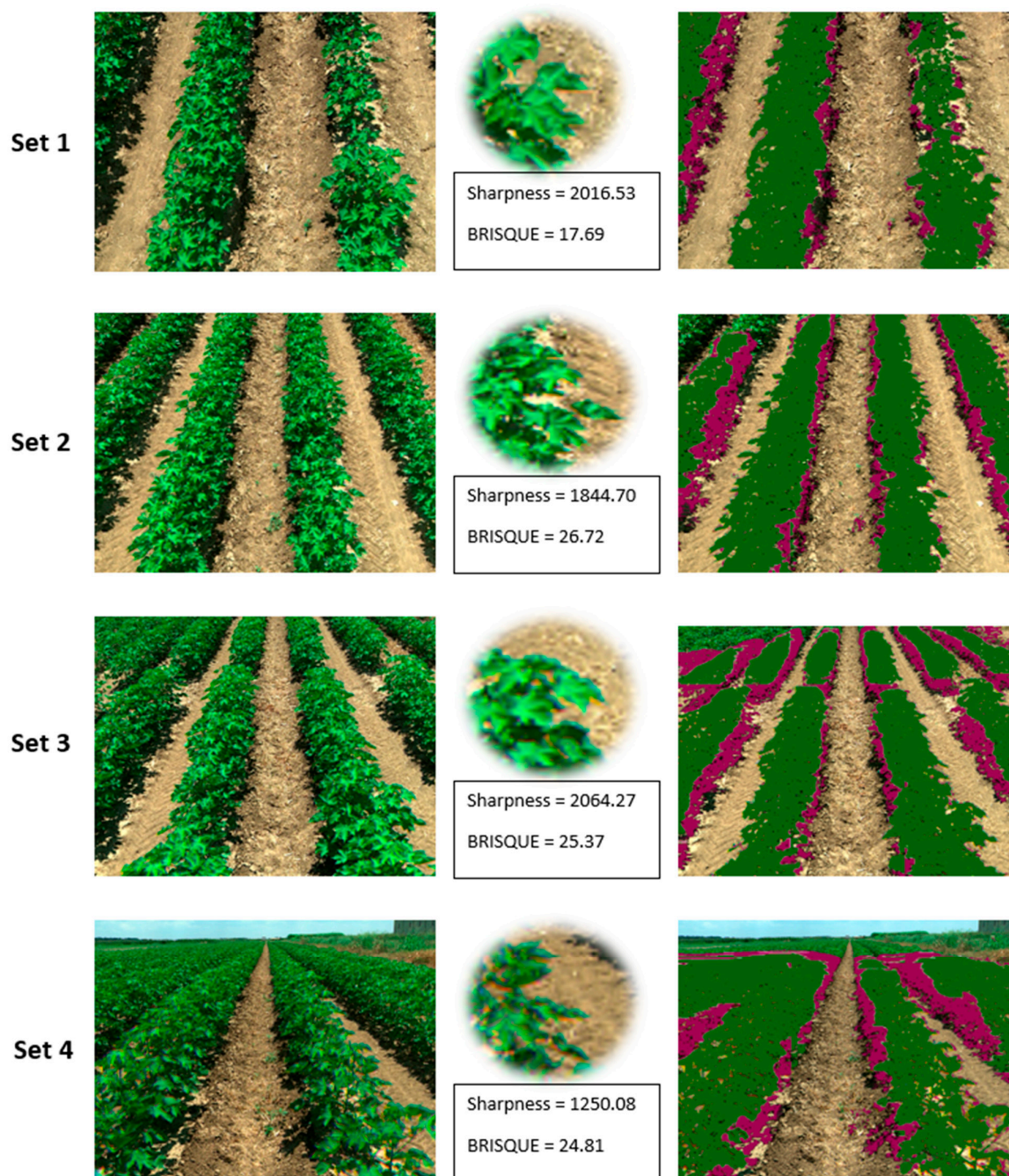


Figure 11. (left column) Aligned RGB images, (middle column) image quality assessment of the RGB images using sharpness and BRISQUE scores, and (right column) vertical segmentation of the cotton canopy layers as upper canopy (UC) (green) and lower canopy (LC) (magenta) for four datasets captured at different camera mounting heights (m) and pitch (°): set 1 (2 m—30°), set 2 (2 m—55°), set 3 (3 m—68°), and set 4 (1.5 m—70°).

3.3. Plant Canopy Segmentation and Spectral Analysis

Figure 11 shows that the vertical segmentation of the plant canopy, as LC and UC, based on their estimated above-ground height, was performed reasonably well by masking out surfaces with zero elevation that represented the soil pixels. However, in some cases, the monocular depth estimation algorithm misidentified shadow pixels on the ground beneath the canopy as plant pixels while estimating above-ground depth, which resulted in the inclusion of a few shadow pixels in the LC layers. The LC layers were more visible in the datasets acquired with the higher camera pitch (set 4). The differences in averaged VI values in the UC and LC layers are shown in Figure 12 for CI_{green} (top) and $CI_{rededge}$ (bottom). While the differences in VIs between the UC and LC layers of set 1 images were not evidently different for the six treatments, they became more evident with increasing camera pitch. In sets 2 and 3, the difference between UC and LC VIs in the no N treatments (T1) was larger in magnitude than in the plots that received N. The low N treatment, T4, also had relatively larger differences in VIs between UC and LC layers, but the other low N treatment, T2, had a response similar to T3, which was subjected to an optimal N rate. However, in set 4, the differences in VIs between the UC and LC layers were negative for the higher N treatments. Since the N treatments did not show the intended response, possibly from movement of nutrients from neighboring plots, VIs from the different canopy layers were compared to the ground-truth leaf N % measurements. A linear regression was performed on the VIs to estimate leaf N %. The comparison of estimated leaf N % with ground-truth measurements showed moderate correlation ($r = 0.53$ to 0.68) for both VIs obtained from the LC and UC layers for sets 1, 3, and 4 (Figures 13 and 14). Also, N % estimated from the differences in VIs between the UC and LC layers showed moderate correlation in sets 1 and 2. However, the RMSE in leaf N % estimation was similar for all cases in Figures 13 and 14, indicating the ambiguity in assessing the accuracy of the different estimation methods.

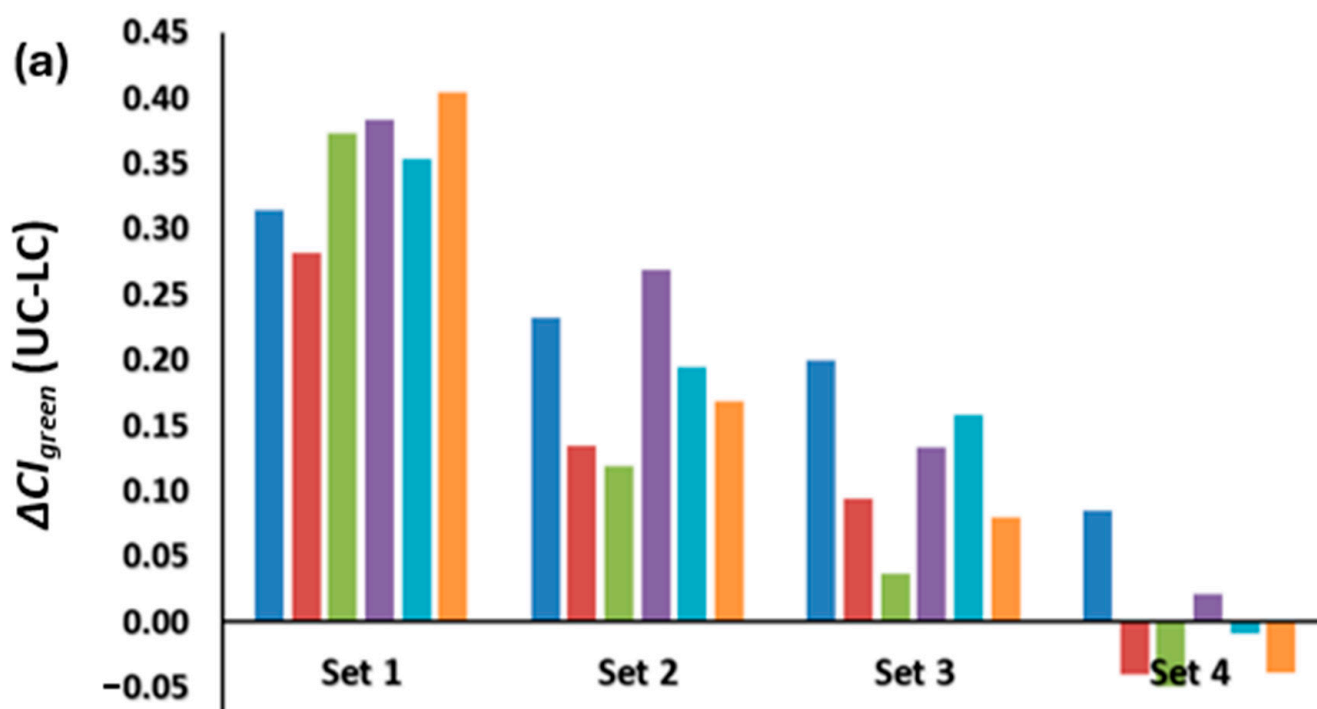


Figure 12. Cont.

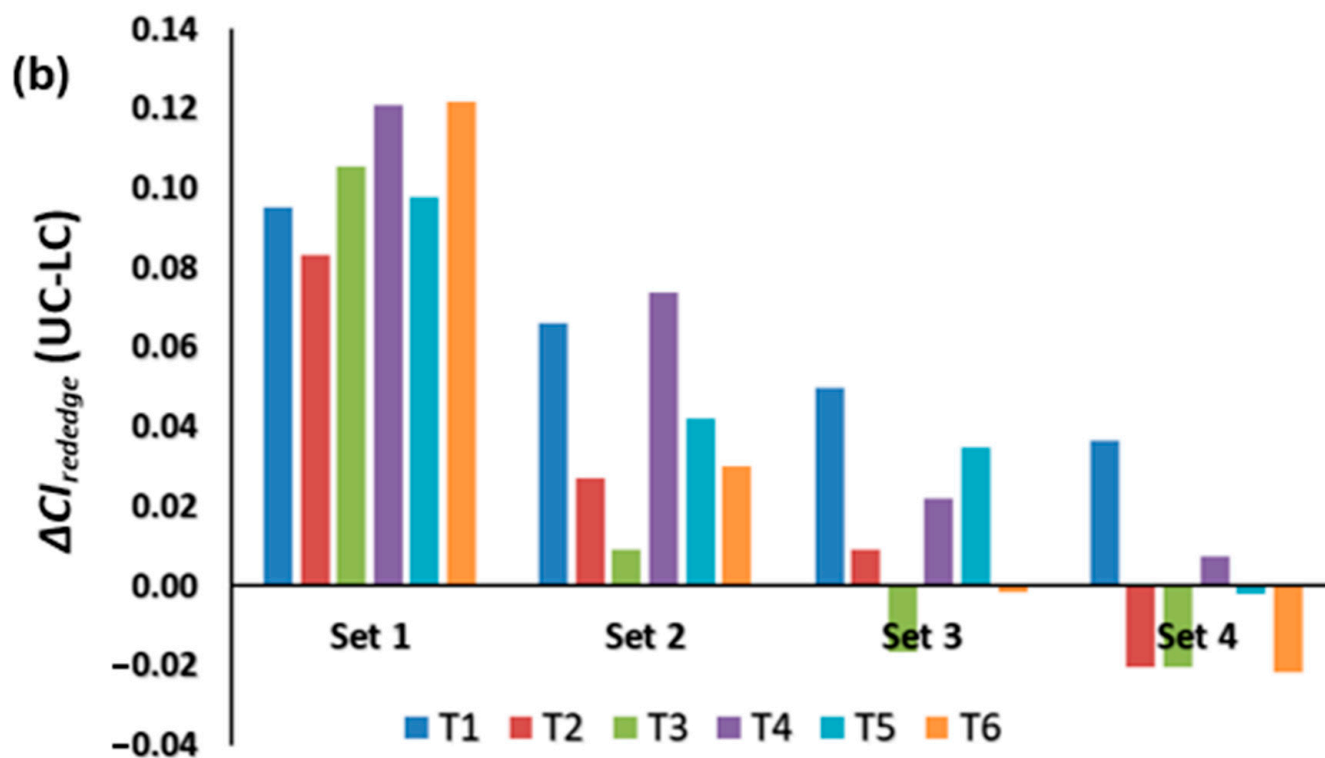


Figure 12. Differences in (a) CI_{green} and (b) $CI_{rededge}$ vegetation indices in the upper canopy (UC) and lower canopy (LC) layers of cotton for images captured at different above-ground mounting heights (m) and camera pitch ($^{\circ}$): set 1 (2 m— 30°), set 2 (3 m— 55°), set 3 (3 m— 68°), and set 4 (1.5 m— 70°). The nitrogen application rate classification for the different treatments are as follows: T1—null, T2 and T4—Low, T3 and T5—optimal, and T6—high.

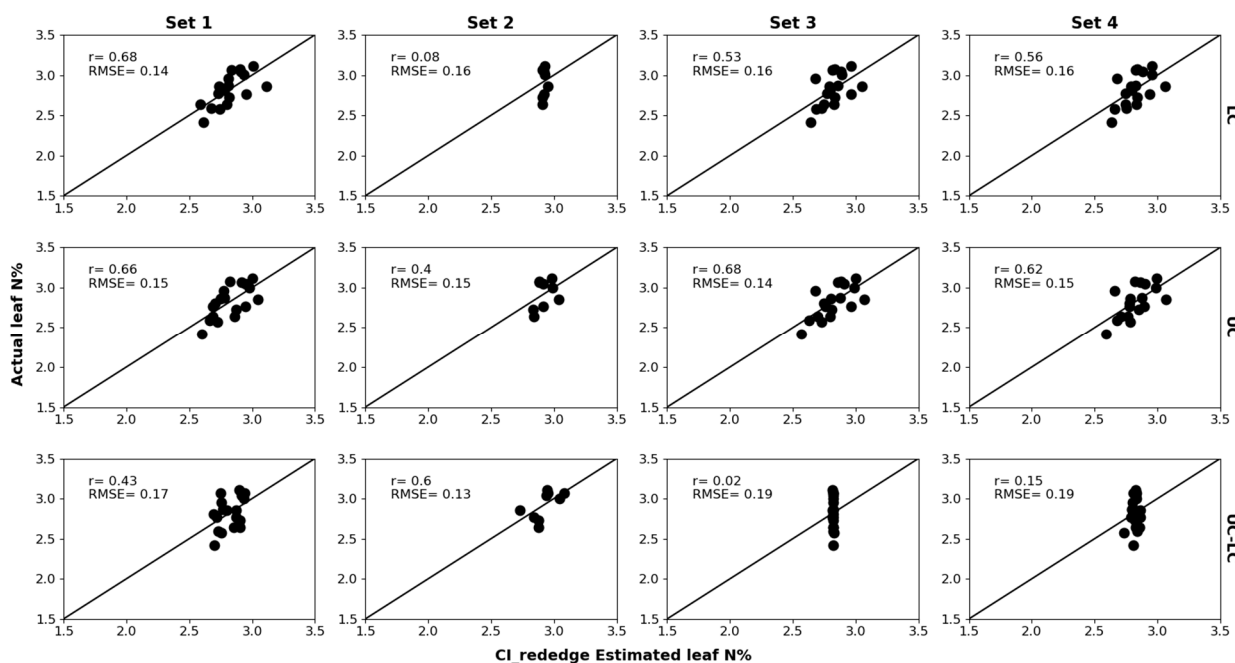


Figure 13. Relationship between actual leaf N % and N % estimate from $CI_{rededge}$ vegetation index obtained from the lower canopy (LC), upper canopy (UC), and differences in UC and LC layers for set 1 (2 m— 30°), set 2 (3 m— 55°), set 3 (3 m— 68°), and set 4 (1.5 m— 70°).

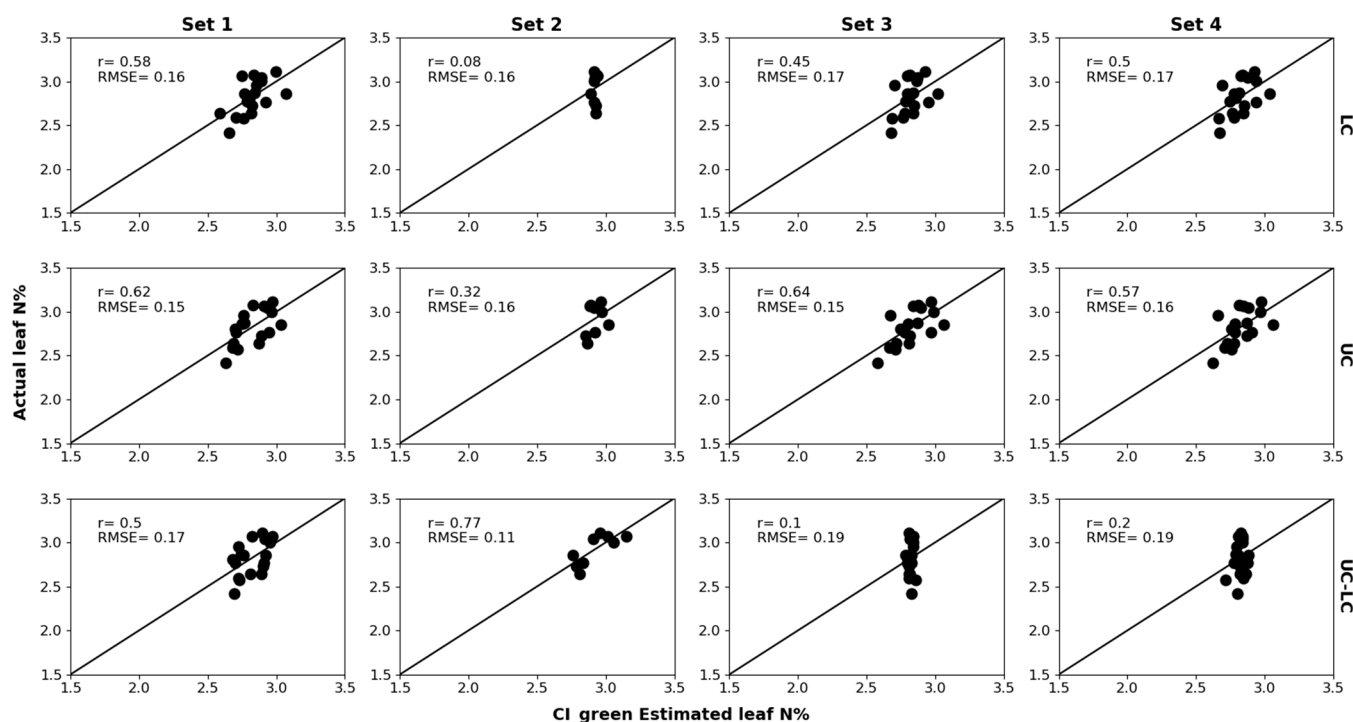


Figure 14. Relationship between actual leaf N % and N % estimate from CI_{green} vegetation index obtained from the lower canopy (LC), upper canopy (UC), and differences in UC and LC layers for set 1 (2 m—30°), set 2 (3 m—55°), set 3 (3 m—68°), and set 4 (1.5 m—70°).

4. Discussion

The use of multispectral cameras to determine plant health status has increased in recent years due to their sensitive of narrow spectral bands, including those beyond the visible light spectrum, and ability to extract morphological plant features. However, multispectral cameras and the software to extract spectral and 3D features from the images have been designed specifically for UAVs or aerial platforms at mid to high image capturing altitudes. We tested the Micasense RedEdge-MX multispectral camera on a ground-based platform for proximal sensing of features that can be potential indicators of early symptoms of N stress in cotton. This study explored the feasibility of using oblique multispectral images for monitoring spectral response in parts of the plant canopy that are generally hidden in conventional nadir-looking images. Compared to multi-sensor ground platforms, this approach aimed at minimizing the vehicle payload, sensor integration and cost of data collection. The camera mount setup used in this study can be adapted to a variety of ground platforms and field conditions. Spectral characteristics of the lower canopy are critical for N stress detection as cotton is known to sustain photosynthesis by remobilizing nutrients from leaves at the bottom to leaves at the top when stressed for N. VIs such as CI_{green} and $CI_{rededge}$ have been determined to be chlorophyll-sensitive, so a different response was expected from the top and bottom canopy layers depending on whether the plants were stressed for N nutrient. The experiment was carried out during squaring, when N uptake and remobilization started increasing rapidly. As the plant canopy grew thicker in the later weeks, it was difficult for the vehicle to pass through the rows and obtain images with sufficient lower canopy details.

A major challenge with the oblique images was tackling the distortion in perspective as objects closer to the camera appeared larger than those farther away from the camera. The perspective effect was especially noticeable in monocular depth estimation, which resulted in an exponential decrease in above-ground canopy height as the distance between the camera and objects increased. Very few studies in the literature have focused on perspective

correction for accurate depth estimation or 3D reconstruction. Liu et al. [42] developed a model for perspective correction with respect to the surface plane, specifically for epipolar rectification between stereo-image pairs. They minimized perspective distortion using the tangent of the angle between the incident light rays and the rectified image plane. In this study, the tangent function of the relative camera pitch (θ_i) that was used for perspective correction in depth was derived empirically.

Despite the perspective correction, differences in estimated depth and height were observed between different images. These differences were verified with the height calibration target, whose estimated above-ground height changed based on its distance from the camera and due to the presence of other objects in the vicinity (Figure 7). Also, the estimated heights were more consistent with the ground truth in one direction than in the other direction for very similar camera setups. A possible reason could be the relative orientation of the crop rows with respect to the camera, which can occur due to differences in steering angles as the vehicle enters the crop rows. Illumination and shading can also contribute to directional variations in monocular depth estimation. These observations indicate that the CNN models were good at providing relative depth estimates within an image but could not maintain consistency between images. Hence, monocular depth estimation remains an ill-posed problem where the output is influenced by various environmental and camera-related factors. Brown and Sukkarieh [22] discussed the concept of ‘object scale ambiguity’ in monocular depth estimation, where it is difficult to accurately estimate the height of two similar objects that differ in size if captured in different frames. However, they also suggested that this issue should not pose a major problem as long as all objects within the camera’s view are of a similar scale. Due to these drawbacks, plant height estimates from monocular depth estimation were not as robust as heights estimated with the multiview stereo/structure from motion (SfM) algorithms popularly used in processing UAV images. Also, SfM algorithms are not suitable for proximal imaging, such as the dataset used in this study, because unique and reliable tie points could not be generated between overlapping images for depth estimation due to the lack of spatial variability within the scenes. Nonetheless, this study highlighted a new method with potential for extracting vertical features from oblique images based on a general-purpose monocular depth estimation model.

Since the Micasense camera lenses have fixed focal lengths, blurriness (low variability in the Laplacian transformation) was observed in images captured closer to the ground, due to poor focus. This fact was supported by the higher sharpness scores of images captured at higher mounting heights or lower pitches in sets 1 (2 m, 30°) and 3 (3 m, 68°). The alignment of images from the different bands was also challenging due to perspective distortion. In addition to perspective induced by the oblique viewing angles of the camera, perspective was also caused by the differences in depth of the canopy and soil with respect to the camera. The lack of reliable SIFT features within the canopy for feature matching (Figure 10) and the relatively poor focus on the canopy due to its proximity to the camera compared to the ground surface may have contributed to the poor alignment of canopy pixels. The mean distance errors and BRISQUE scores indicated that the overall image misalignment further increased with increasing pitch and decreasing camera mounting height because of greater distortion in perspective. Although the low oblique images in set 1 provided better alignment and image quality, they did not capture the lower layers of the canopy sufficiently, as required for this study. The segmentation results showed that images with higher pitch that were closer to ground (set 4 > set 3 > set 2 > set 1) were better at capturing the lower layers of the canopy. Thus, the differences in VIs between UC and LC layers appeared similar for all N treatments in set 1, compared to sets 2 and 3, which had higher pitches. In sets 2 and 3, the differences in VIs between the UC and

LC layers were consistently high and low for T1 (no N) and T3 (optimal N), respectively. This supports the assumption that the lower canopy leaves are likely to senesce early in response to N stress. However, the magnitude of differences between the UC and LC VIs for the other plots varied for each set and VI. While this approach may show some distinction between N deficient and sufficient plots, it may not be a good indicator of the stress levels without achieving good image alignment and quality. The negative differences in UC and LC VIs prevalently observed in set 4 was likely due to the high degree of misalignment that resulted in incorrect estimation of VI values in the UC and LC layers. Having lower health-related VI values in the UC than the LC is, in reality, highly unlikely in the early cotton growth stages, as the UC leaves are newer and healthier than the LC leaves and often N translocation happens from the LC to UC layers to maintain photosynthetic efficiency. Also, different VI responses were observed between treatment pairs that had the same application rates (T2 and T4, and T3 and T5), possibly due to differences in application timing, variabilities in the soil conditions, and movement of the fertilizer across the plots due to rainfall. Although leaf N % estimated with VIs extracted from the LC and UC layers showed moderate correlation with ground-truth measurements, more data are required to conclusively determine the response of the different canopy layers to N availability and uptake. Self-shading of the leaves and misclassification of leaf shadows on the ground as canopy pixels also induce errors in VI calculations. Shadow elimination with thresholding and image transformation techniques were ineffective since the images were highly textured due to shadows on the canopy and misalignment. Hence, better ground-truthing and radiometric calibration techniques are needed to verify the observed differences in the upper and lower canopy spectral responses. Validating the spectral data with spectrometer measurements may not be reliable under field conditions due to shading and bidirectional reflectance. However, the spectral response of the different cotton canopy layers in response to N stress can be studied in controlled environments.

In summary, this study explored the possibility of extracting spectral and physical features of cotton plants under field conditions from oblique ground-based multispectral images. In addition to the challenges in obtaining oblique images under field conditions, this study also addressed challenges associated with processing those images. Future work should focus on improving alignment between the different image bands, a process that may include CNN models like the NST [26] and RegiNet [18]. RGB images of crop rows acquired with similar pitch and height settings can be used as references for quality assessments of aligned multispectral images. Further, this study did not account for the changes in exposure time and gain that can affect the radiometric quality of the image and the derived VIs. Auto-exposure settings result in dynamic adjustment of exposure time and gain according to the brightness or reflectance range of objects in the scene [40]. A few studies have determined that the use of fixed exposure settings is more effective in producing radiometric accuracy in UAV images [43,44]. These methods should be explored for ground-based images, too, in order to obtain more reliable estimates of VIs from the different canopy layers that are subjected to different degrees of shading. Developing stereovision techniques for multi-lens cameras like the Micasense camera series or training monocular depth estimation models on custom row crop datasets can also help improve 2D image-based height estimation accuracy.

5. Summary and Conclusions

This study shows the potential of oblique ground-based imaging in deriving useful information from the lower parts of the cotton plant canopy, which are generally hidden from view in nadir-looking images. Methods were established to acquire, process, and extract quantitative features from oblique ground-based images. The major contributions of

this study are (1) the development of a technique to correct perspective-induced distortion in monocular depth and height estimation, and (2) a spectral analysis of different cotton canopy layers in response to N availability and uptake. The results showed promise in the extraction of physical plant characteristics such as plant height from oblique images. While the spectral analysis with vegetation indices indicates the possibility of differential response in the upper and lower canopy layers due to N stress, it was difficult to conclude whether this approach will be beneficial in identifying early symptoms of N stress in field cotton. Further work with extensive data collection is needed to perfect the methods in this new approach toward quantitative feature extraction from proximally sensed oblique images.

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

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: We acknowledge the assistance received from Roy Graves, Amrit Shreshta, Karem Meza Capcha and the undergraduate student workers with field data collection.

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

References

1. Alexopoulos, A.; Koutras, K.; Ali, S.B.; Puccio, S.; Carella, A.; Ottaviano, R.; Kalogeras, A. Complementary use of ground-based proximal sensing and airborne/spaceborne remote sensing techniques in precision agriculture: A systematic review. *Agronomy* **2023**, *13*, 1942. [\[CrossRef\]](#)
2. Sui, R.; Thomasson, J.A. Ground-based sensing system for cotton nitrogen status determination. *Trans. ASABE* **2006**, *49*, 1983–1991. [\[CrossRef\]](#)
3. Jin, X.; Zarco-Tejada, P.J.; Schmidhalter, U.; Reynolds, M.P.; Hawkesford, M.J.; Varshney, R.K.; Yang, T.; Nie, C.; Li, Z.; Ming, B.; et al. High-throughput estimation of crop traits: A review of ground and aerial phenotyping platforms. *IEEE Geosci. Remote Sens. Mag.* **2020**, *9*, 200–231. [\[CrossRef\]](#)
4. Mueller-Sim, T.; Jenkins, M.; Abel, J.; Kantor, G. The Robotanist: A ground-based agricultural robot for high-throughput crop phenotyping. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; IEEE: Piscataway, NJ, USA, 2017.
5. Jiang, Y.; Li, C.; Robertson, J.S.; Sun, S.; Xu, R.; Paterson, A.H. GPhenoVision: A ground mobile system with multi-modal imaging for field-based high throughput phenotyping of cotton. *Sci. Rep.* **2018**, *8*, 1213. [\[CrossRef\]](#)
6. Issac, A.; Ebrahimi, A.; Velni, J.M.; Rains, G. Development and Deployment of a Big Data Pipeline for Field-based High-throughput Cotton Phenotyping Data. *Smart Agric. Technol.* **2023**, *5*, 100265. [\[CrossRef\]](#)
7. Antonucci, F.; Costa, C.; Figorilli, S.; Ortenzi, L.; Manganiello, R.; Santangelo, E.; Gierz, Ł.; Pallottino, F. A Low-Cost Sensorized Vehicle for In-Field Crop Phenotyping. *Appl. Sci.* **2023**, *13*, 2436. [\[CrossRef\]](#)
8. Hu, C.; Xie, S.; Song, D.; Thomasson, J.A.; Hardin, I.V.R.G.; Bagavathiannan, M. Algorithm and system development for robotic micro-volume herbicide spray towards precision weed management. *IEEE Robot. Autom. Lett.* **2022**, *7*, 11633–11640. [\[CrossRef\]](#)
9. Xiong, Y.; Ge, Y.; Grimstad, L.; From, P.J. An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation. *J. Field Robot.* **2020**, *37*, 202–224. [\[CrossRef\]](#)
10. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; IEEE: Piscataway, NJ, USA, 1999.
11. Karami, E.; Prasad, S.; Shehata, M. Image matching using SIFT, SURF, BRIEF and ORB: Performance comparison for distorted images. *arXiv* **2017**, arXiv:1710.02726.
12. Tareen, S.A.K.; Saleem, Z. A comparative analysis of sift, surf, kaze, akaze, orb, and brisk. In Proceedings of the 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 3–4 March 2018; IEEE: Piscataway, NJ, USA, 2018.

13. Dong, J.; Burnham, J.G.; Boots, B.; Rains, G.; Dellaert, F. 4D crop monitoring: Spatio-temporal reconstruction for agriculture. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; IEEE: Piscataway, NJ, USA, 2017.
14. Islam, R.; Habibullah, H.; Hossain, T. AGRI-SLAM: A real-time stereo visual SLAM for agricultural environment. *Auton. Robot.* **2023**, *47*, 649–668. [\[CrossRef\]](#)
15. Yun, C.; Kim, H.; Jeon, C.; Gang, M.; Lee, W.S.; Han, J.G. Stereovision-based ridge-furrow detection and tracking for auto-guided cultivator. *Comput. Electron. Agric.* **2021**, *191*, 106490. [\[CrossRef\]](#)
16. Kim, W.; Lee, D.; Kim, Y.; Kim, T.; Lee, W.; Choi, C. Stereo-vision-based crop height estimation for agricultural robots. *Comput. Electron. Agric.* **2021**, *181*, 105937. [\[CrossRef\]](#)
17. Bao, Y.; Tang, L.; Breitzman, M.W.; Salas Fernandez, M.G.; Schnable, P.S. Field-based robotic phenotyping of sorghum plant architecture using stereo vision. *J. Field Robot.* **2019**, *36*, 397–415. [\[CrossRef\]](#)
18. Wei, Z.; Jung, C.; Su, C. RegiNet: Gradient guided multispectral image registration using convolutional neural networks. *Neurocomputing* **2020**, *415*, 193–200. [\[CrossRef\]](#)
19. da Silva, M.V.; Silva, L.H.; Junior, J.D.D.; Escarpinati, M.C.; Backes, A.R.; Mari, J.F. Generating synthetic multispectral images using neural style transfer: A study with application in channel alignment. *Comput. Electron. Agric.* **2023**, *206*, 107668. [\[CrossRef\]](#)
20. Olenskyj, A.G.; Sams, B.S.; Fei, Z.; Singh, V.; Raja, P.V.; Bornhorst, G.M.; Earles, J.M. End-to-end deep learning for directly estimating grape yield from ground-based imagery. *Comput. Electron. Agric.* **2022**, *198*, 107081. [\[CrossRef\]](#)
21. Huang, Y.; Lee, M.A.; Thomson, S.J.; Reddy, K.N. Ground-based hyperspectral remote sensing for weed management in crop production. *Int. J. Agric. Biol. Eng.* **2016**, *9*, 98–109.
22. Brown, J.; Sukkarieh, S. Improving monocular depth prediction in ambiguous scenes using a single range measurement. *IFAC-PapersOnLine* **2019**, *52*, 355–360. [\[CrossRef\]](#)
23. Gutiérrez, S.; Wendel, A.; Underwood, J. Ground based hyperspectral imaging for extensive mango yield estimation. *Comput. Electron. Agric.* **2019**, *157*, 126–135. [\[CrossRef\]](#)
24. Tanaka, Y.; Watanabe, T.; Katsura, K.; Tsujimoto, Y.; Takai, T.; Tanaka, T.S.T.; Kawamura, K.; Saito, H.; Homma, K.; Mairoua, S.G.; et al. Deep learning enables instant and versatile estimation of rice yield using ground-based RGB images. *Plant Phenomics* **2023**, *5*, 0073. [\[CrossRef\]](#)
25. Sethy, P.K.; Barpanda, N.K.; Rath, A.K.; Behera, S.K. Nitrogen deficiency prediction of rice crop based on convolutional neural network. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *11*, 5703–5711. [\[CrossRef\]](#)
26. Zhao, G.; Cai, W.; Wang, Z.; Wu, H.; Peng, Y.; Cheng, L. Phenotypic parameters estimation of plants using deep learning-based 3-D reconstruction from single RGB image. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [\[CrossRef\]](#)
27. Cui, X.; Feng, Q.; Wang, S.; Zhang, J. Monocular depth estimation with self-supervised learning for vineyard unmanned agricultural vehicle. *Sensors* **2022**, *22*, 721. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Ranftl, R.; Bochkovskiy, A.; Koltun, V. Vision transformers for dense prediction. *arXiv* **2021**, arXiv:2103.13413.
29. Ranftl, R.; Lasinger, K.; Hafner, D.; Schindler, K.; Koltun, V. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 1623–1637. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Zhao, C.; Zhang, Y.; Poggi, M.; Tosi, F.; Guo, X.; Zhu, Z.; Huang, G.; Tang, Y.; Mattoccia, S. Monovit: Self-supervised monocular depth estimation with a vision transformer. In Proceedings of the 2022 International Conference on 3D Vision (3DV), Prague, Czech Republic, 12–16 September 2022; IEEE: Piscataway, NJ, USA, 2022.
31. Tangellapalli, S.A.; Sangha, H.S.; Peschel, J.; Duncan, B.A. Assessing Monocular Depth Estimation Networks for UAS Deployment in Rainforest Environments. In Proceedings of the 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Abu Dhabi, United Arab Emirates, 14–18 October 2024; IEEE: Piscataway, NJ, USA, 2024.
32. Mizuta, K.; Sato, Y.; Hayashi, J.; Toyota, M.; Morokuma, M. Canopy coverage of wheat measured with high accuracy using the HSV colour model and relative depth estimation model, MiDaS. *Plant Prod. Sci.* **2024**, *27*, 294–303. [\[CrossRef\]](#)
33. Gharakhani, H.; Thomasson, J.A.; Nematzadeh, P.; Yadav, P.K.; Hague, S. Using under-canopy cotton imagery for cotton variety classification. In Proceedings of the Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping VII, Orlando, FL, USA, 3–7 April 2022; SPIE: Philadelphia, PA, USA, 2022.
34. Xiao, S.; Ye, Y.; Fei, S.; Chen, H.; Cai, Z.; Che, Y.; Wang, Q.; Ghafoor, A.; Bi, K.; Shao, K.; et al. High-throughput calculation of organ-scale traits with reconstructed accurate 3D canopy structures using a UAV RGB camera with an advanced cross-circling oblique route. *ISPRS J. Photogramm. Remote Sens.* **2023**, *201*, 104–122. [\[CrossRef\]](#)
35. Zhu, B.; Zhang, Y.; Sun, Y.; Shi, Y.; Ma, Y.; Guo, Y. Quantitative estimation of organ-scale phenotypic parameters of field crops through 3D modeling using extremely low altitude UAV images. *Comput. Electron. Agric.* **2023**, *210*, 107910. [\[CrossRef\]](#)
36. Che, Y.; Wang, Q.; Xie, Z.; Zhou, L.; Li, S.; Hui, F.; Wang, X.; Li, B.; Ma, Y. Estimation of maize plant height and leaf area index dynamics using an unmanned aerial vehicle with oblique and nadir photography. *Ann. Bot.* **2020**, *126*, 765–773. [\[CrossRef\]](#)
37. Li, M.; Shamshiri, R.R.; Schirrmann, M.; Weltzien, C. Impact of camera viewing angle for estimating leaf parameters of wheat plants from 3D point clouds. *Agriculture* **2021**, *11*, 563. [\[CrossRef\]](#)

38. Hasan, M.M.; Chopin, J.P.; Laga, H.; Miklavcic, S.J. Detection and analysis of wheat spikes using convolutional neural networks. *Plant Methods* **2018**, *14*, 100. [[CrossRef](#)]
39. Nelson, D.W.; Sommers, L.E. Determination of total nitrogen in plant material. *Agron. J.* **1973**, *65*, 109–112. [[CrossRef](#)]
40. Evangelidis, G.D.; Psarakis, E.Z. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1858–1865. [[CrossRef](#)] [[PubMed](#)]
41. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [[CrossRef](#)] [[PubMed](#)]
42. Liu, J.; Guo, B.; Jiang, W.; Gong, W.; Xiao, X. Epipolar rectification with minimum perspective distortion for oblique images. *Sensors* **2016**, *16*, 1870. [[CrossRef](#)] [[PubMed](#)]
43. Swaminathan, V.; Thomasson, J.A.; Hardin, R.G.; Rajan, N. Selection of appropriate multispectral camera exposure settings and radiometric calibration methods for applications in phenotyping and precision agriculture. *arXiv* **2024**, arXiv:2402.18553. [[CrossRef](#)]
44. Bagnall, G.C.; Thomasson, J.A.; Yang, C.; Wang, T.; Han, X.; Sima, C.; Chang, A. Uncrewed aerial vehicle radiometric calibration: A comparison of autoexposure and fixed-exposure images. *Plant Phenome J.* **2023**, *6*, e20082. [[CrossRef](#)]

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