agriculture

MDPI

Article

A Novel Approach to Optimize Key Limitations of Azure Kinect DK for Efficient and Precise Leaf Area Measurement

Ziang Niu [1], Ting Huang [1], Chengjia Xu [1], Xinyue Sun [1], Mohamed Farag Taha [2,3], Yong He [1] and Zhengjun Qiu [1,*]

[1] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China; niu_ziang@zju.edu.cn (Z.N.); t3huang@zju.edu.cn (T.H.); 22213017@zju.edu.cn (C.X.); 22313020@zju.edu.cn (X.S.); yhe@zju.edu.cn (Y.H.)
[2] School of Agricultural Engineering, Jiangsu University, Zhenjiang 212013, China; 1000006483@ujs.edu.cn
[3] Department of Soil and Water Sciences, Faculty of Environmental Agricultural Sciences, Arish University, North Sinai 45516, Egypt
* Correspondence: zjqiu@zju.edu.cn

Abstract: Maize leaf area offers valuable insights into physiological processes, playing a critical role in breeding and guiding agricultural practices. The Azure Kinect DK possesses the real-time capability to capture and analyze the spatial structural features of crops. However, its further application in maize leaf area measurement is constrained by RGB–depth misalignment and limited sensitivity to detailed organ-level features. This study proposed a novel approach to address and optimize the limitations of the Azure Kinect DK through the multimodal coupling of RGB-D data for enhanced organ-level crop phenotyping. To correct RGB–depth misalignment, a unified recalibration method was developed to ensure accurate alignment between RGB and depth data. Furthermore, a semantic information-guided depth inpainting method was proposed, designed to repair void and flying pixels commonly observed in Azure Kinect DK outputs. The semantic information was extracted using a joint YOLOv11-SAM2 model, which utilizes supervised object recognition prompts and advanced visual large models to achieve precise RGB image semantic parsing with minimal manual input. An efficient pixel filter-based depth inpainting algorithm was then designed to inpaint void and flying pixels and restore consistent, high-confidence depth values within semantic regions. A validation of this approach through leaf area measurements in practical maize field applications—challenged by a limited workspace, constrained viewpoints, and environmental variability—demonstrated near-laboratory precision, achieving an MAPE of 6.549%, RMSE of 4.114 cm^2, MAE of 2.980 cm^2, and R^2 of 0.976 across 60 maize leaf samples. By focusing processing efforts on the image level rather than directly on 3D point clouds, this approach markedly enhanced both efficiency and accuracy with the sufficient utilization of the Azure Kinect DK, making it a promising solution for high-throughput 3D crop phenotyping.

Keywords: 3D; phenotyping; misalignment; depth quality; inpainting

check for updates

Academic Editor: Josef Eitzinger

Received: 17 December 2024
Revised: 8 January 2025
Accepted: 11 January 2025
Published: 14 January 2025

Citation: Niu, Z.; Huang, T.; Xu, C.; Sun, X.; Taha, M.F.; He, Y.; Qiu, Z. A Novel Approach to Optimize Key Limitations of Azure Kinect DK for Efficient and Precise Leaf Area Measurement. *Agriculture* **2025**, *15*, 173. https://doi.org/10.3390/agriculture15020173

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

The leaf area of maize is intricately linked to key physiological processes such as photosynthesis, above-ground net primary productivity, and dry matter accumulation, serving as a crucial indicator of how the plant responds to factors like variety, environmental conditions, and cultivation practices. By revealing the plant's underlying growth and developmental patterns across varying environments, this parameter holds significant

potential in advancing maize breeding programs and facilitating research in functional–structural plant models (FSPMs) [1,2].

Traditionally, gathering morphological data relies heavily on manual measurement, which is labor-intensive, time-consuming, and susceptible to errors due to varying measurement tools and operator differences [3,4]. With advancements in sensor technology and computer vision, high-throughput, high-precision, non-destructive measurement techniques are increasingly applied in crop phenotyping. Notably, three-dimensional (3D) phenotyping technologies have emerged as a research hotspot, offering comprehensive analyses of crop morphological structures in spatial dimensions. Compared to image-based phenotyping methods [5,6], 3D techniques provide true scale measurements and allow for a more detailed analysis of the topological structure of the crop. It is worth noting that while larger-scale sensing technologies excel at capturing extensive crop canopy data, their limitations in resolution and organ-level detail restrict their utility in high-precision phenotyping tasks, such as measuring maize leaf area. Therefore, this study focuses on leveraging proximal sensing technologies for detailed maize leaf area analysis. To address its restricted coverage area, integrating the device with a mobile platform or multi-device setup can enable high-throughput data acquisition while maintaining measurement accuracy.

A comparison of different 3D sensing technologies and devices is shown in Table 1. Technologies such as LiDAR scanners [7–9], structured light cameras, and structure from motion (SfM) methods [10,11] have diverse applications in agricultural contexts. Among these, RGB-D cameras stand out for their instant imaging capability, low cost, and portability, making them particularly advantageous for in-field, high-throughput phenotyping [12]. The Kinect series, designed by Microsoft, has gained significant popularity, with its three generations of consumer-grade cameras—Kinect V1, Kinect V2, and the Azure Kinect DK—widely reported and applied in agricultural practices. The first-generation Kinect employed environmentally sensitive structured light for depth measurement, offering a relatively low resolution suitable for rough estimations of objects. For instance, Andújar et al. used Kinect V1 to estimate the biomass of cauliflower crops [13]. Kinect V2, which adopted time-of-flight (ToF) technology, achieved improved depth estimations. The device demonstrated strong sensing capabilities for canopies of leafy crops or seedlings when combined with multi-view data [14–17] and even enabled detailed analyses of crop structures such as plant skeletons or leaves [18,19]. Published in 2019, it has been reported that the latest generation, the Azure Kinect DK, surpasses other consumer-grade RGB-D cameras in terms of system depth accuracy and resolution quality [20], demonstrating a superior performance in dynamic environments [21] and semi-outdoor scenarios [22].

Table 1. The comparison of different 3D sensing technologies and devices.

Indicator	LiDAR Scanner	Structured Light Camera	SfM	RGB-D Camera (Azure Kinect DK)
Sensing method	Active	Active	Passive	Active
Resolution	High	Medium	Medium	Medium
Accuracy	High	High	Medium	Medium
Environmental robustness	High	Low	Low	Medium
Real-time performance	Low	Medium	Low	High
cost	High	High	Low	Medium

Despite the widespread utilization of the Azure Kinect DK, limitations stemming from hardware constraints and condition variabilities restrict its application for organ-level tasks like leaf area measurement. Though leaf area measurement for detached single leaves can achieve an RMSE of 4.05 cm^2 [23] in laboratory settings, measurements for the

whole maize plants in the field only achieved an RMSE ranging from 11 to 41 cm^2 [24,25]. The results suggest that the inherent sensing limitations of the RGB-D camera necessitate the application of supplementary algorithms to improve measurement precision, thereby enhancing its practical utility [26]. Although researchers have proposed methods such as point cloud optimization [27,28] and occlusion completion [29,30] to improve data accuracy, these approaches lack targeted solutions that address the specific challenges posed by the Azure Kinect DK in leaf area measurement, as discussed below.

There are three significant challenges of the Azure Kinect DK for crop phenotyping, especially for leaf area measurement. First, the misalignment between RGB and depth data constrains the utility of visual data and limits the application of RGB-derived semantic information in subsequent analyses. Second, the quality of depth data is not adequate for the detailed perception of small structures and edges, which frequently results in void pixels. Third, the Azure Kinect DK lacks clear delineation along object edges, where the presence of flying pixels severely impacts the accuracy for applications relying on spatial structure analysis. Therefore, the objectives of this study are as follows: (1) to propose a unified recalibration method for correcting the misalignment between the RGB and depth data of the Azure Kinect DK; (2) to leverage the semantic characteristics of RGB data to identify void pixels and flying pixels within the leaf regions of depth data that affect 3D phenotyping; (3) to design a depth inpainting process to repair depth information for these critical pixels; and (4) to validate the proposed approach in practical maize field applications, in which limited workspaces, constrained viewpoints, and unstructured variations in wind speed and lighting conditions pose additional challenges to leaf area measurement, thereby optimizing the key limitations of the Azure Kinect DK. These optimizations hold significant potential for advancing agricultural precision management. By enhancing the accuracy and reliability of leaf area measurement, the proposed method facilitates more the precise monitoring of crop growth and health, enabling informed decision making for irrigation, fertilization, and pest control. Additionally, the improved capability of the Azure Kinect DK to handle real-world field challenges allows for the scalable and cost-effective adoption of 3D phenotyping technologies in diverse agricultural scenarios, promoting sustainable and efficient crop management practices.

2. Materials and Methods

2.1. Experiment Setup

The Microsoft Azure Kinect DK (as shown in Figure 1a) is an advanced RGB-D camera based on continuous wave ToF technology, with a reported random error standard deviation of $\leq$17 mm. Additionally, its typical system error is <11 mm +0.1% of the measured distance in the absence of multipath interference. The Azure Kinect DK is equipped with a high-resolution RGB camera with a maximum resolution of 12 MP and a depth sensor with a maximum resolution of 1 MP. Operating at a laser wavelength of 850 nm, the depth sensor determines depth by measuring the phase difference between transmitted and received signals, which enables the simultaneous acquisition of depth images and their corresponding infrared images. The study selected WFOV (Wide Field of View) mode without binning of Azure Kinect DK for closer sensing distance and improved data quality. Under this mode, the FOV of RGB camera is H 90° and V 59°, and the RGB image resolution and frame rate are set to 3840 × 2160 at 5 fps. Meanwhile, the FOV of depth sensor is H120° V120°, with depth image resolution and frame rate set to 1024 × 1024 at 5 fps.

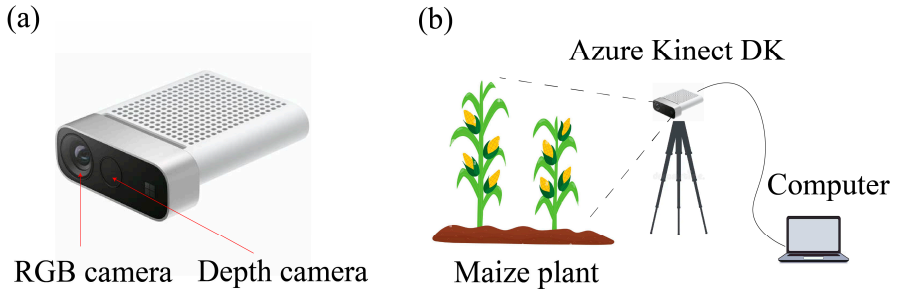


Figure 1. Experiment setup. (**a**) Azure Kinect DK. (**b**) Experiment setup.

The field experiment of this study was conducted at the agricultural experimental field of Zhejiang University, Hangzhou, China. Data of maize plants were collected in August 2024, including RGB images, depth images, and ground truth of leaf areas. To meet the demand for high-throughput and rapid phenotyping in the field, this study employed a single-view approach using Azure Kinect DK while ensuring the maximum exposure of the entire leaf area within the field of view. In the experiment, the RGB-D camera was connected to a mobile workstation via a cable and mounted on a tripod or handheld, as shown in Figure 1b. The data were collected and saved by the program developed based on the Azure Kinect SDK 1.4.

Fifteen maize plants with total 60 leaf samples were carefully selected to ensure the dataset represented the diversity of maize crops in real-world agricultural settings. The selection criteria included the following:

- Growth Stage Diversity: Samples were chosen from various growth stages, ranging from V1 (vegetative stage when the first leaf collar is present) to V7 (vegetative stage when seven leaf collars are present);
- Plant Size Variability: Maize plants with varying heights (from ~20 cm to ~45 cm) and leaf sizes (from ~10 cm^2 to ~100 cm^2) were included to reflect natural variability;
- Spatial Distribution: Plants from different locations within the field were selected to account for variations caused by soil fertility, shading, and other factors.

During data collection, environmental conditions were monitored to ensure consistency and to simulate real-world field challenges. Ambient temperatures ranged from 25 °C to 35 °C, and wind speeds averaged 2–5 m/s. To minimize the impact of environmental variables, data collection was performed in early morning or late afternoon, when sunlight was softer, reducing glare and shadow effects in the images.

The ground truth of leaf areas was obtained by cutting each maize leaf into separate parts, which were then immediately flattened and scanned using a 2D scanner. Additionally, 100 RGB images of maize plants were collected to train the semantic information extraction network as discussed in Section 2.3. All of the tasks in this study were conducted on a hardware platform equipped with an AMD Ryzen 7 3700X 8-Core CPU Processor, 32 GB of memory, and an NVIDIA GTX 1080Ti GPU.

2.2. Problem Description

Following the experiment setup illustrated in the Section 2.1, the RGB-D data generated by the Azure Kinect DK presents three main challenges that impact subsequent processing and applications. First, since the RGB camera and depth sensor are heterogeneous cameras integrated in the Azure Kinect DK, their data require alignment for further applications. However, low alignment accuracy has been found when using pre-shipment calibration parameters, especially for close-range objects [31]. As shown in Figure 2a, green regions that originally belonged to the leaf were misprojected onto the ground. Second, due to the sensing capability limitation of Azure Kinect DK, the depth data cannot represent objects

with surface reflections, occlusions, or out of measuring range, resulting in void pixels [32], as shown in Figure 2b, in which the blue line indicates the target region. Third, along the edges of object, depth measurements are affected by mixed reflections from both the object and the background, resulting in a reduced signal-to-noise ratio. This interference produces cascading "flying pixels" between the object and background [33,34], as illustrated in Figure 2c. To address these issues, the following methods were employed to mitigate their effects.

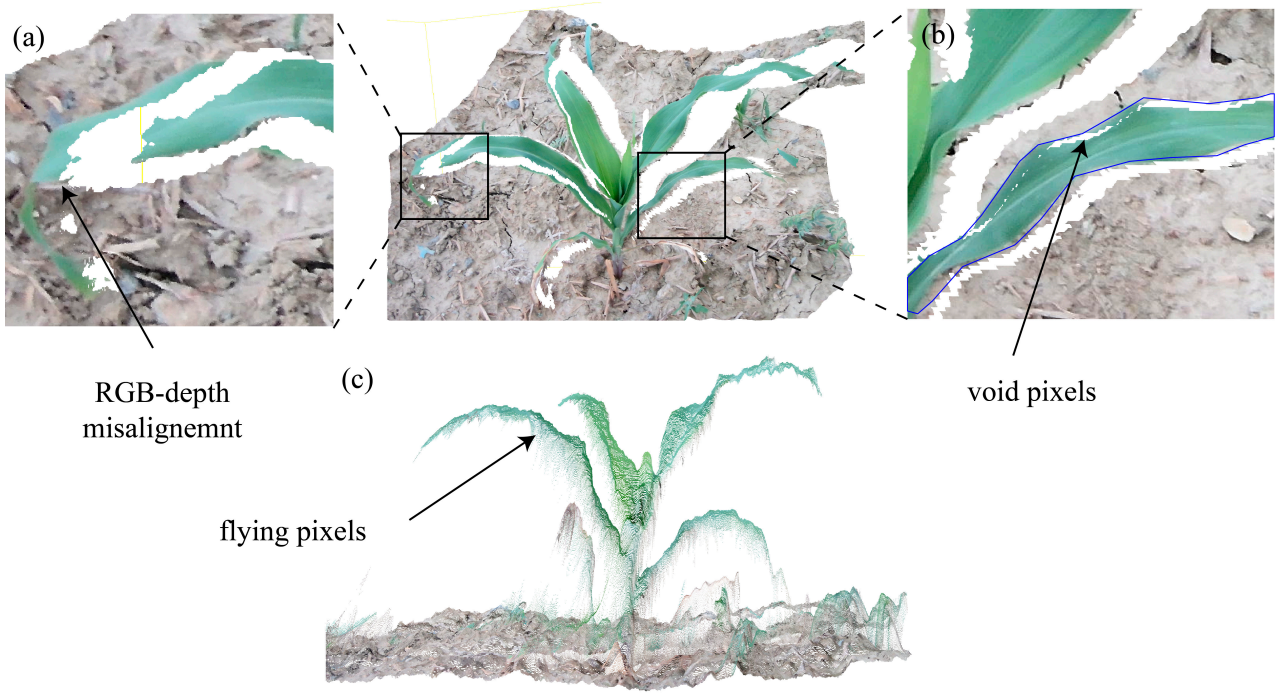


Figure 2. Key limitations of the data generated by Azure Kinect DK. (**a**) Front-view point cloud of Azure Kinect DK showing RGB–depth misalignment. (**b**) Front-view point cloud showing void pixels inside leaf region. (**c**) Side-view point cloud showing flying pixels around the edges of leaf.

2.3. Recalibration of RGB-D Camera

Since the algorithm relied on precise alignment between the RGB and depth images, this study conducted a recalibration method based on checkerboard for the extrinsic parameters of RGB and depth cameras. As introduced in Section 2.1, although the depth sensor lacks sensitivity to optical data, its raw infrared images can be utilized for detecting checkerboard corner points [35]. A black-and-white checkerboard with dimensions of 15 cm × 10 cm and a 9 × 6 grid pattern was employed as the calibration target. Positioned within a 20 cm to 1 m range from the RGB-D camera, the checkerboard was moved and rotated across various field-of-view positions, yielding a dataset of 20 RGB and infrared image pairs. The calibration was performed using MATLAB's Stereo Camera Calibrator tool to extract the extrinsic parameters for both the RGB and depth cameras. The depth image was aligned to the RGB image through a subsequent coordinate transformation [36], as illustrated in Figure 3. It is worth noting that the recalibration only needs to be performed once. Subsequent applications can use the extrinsic parameters obtained from this recalibration to produce consistently aligned RGB–depth results.

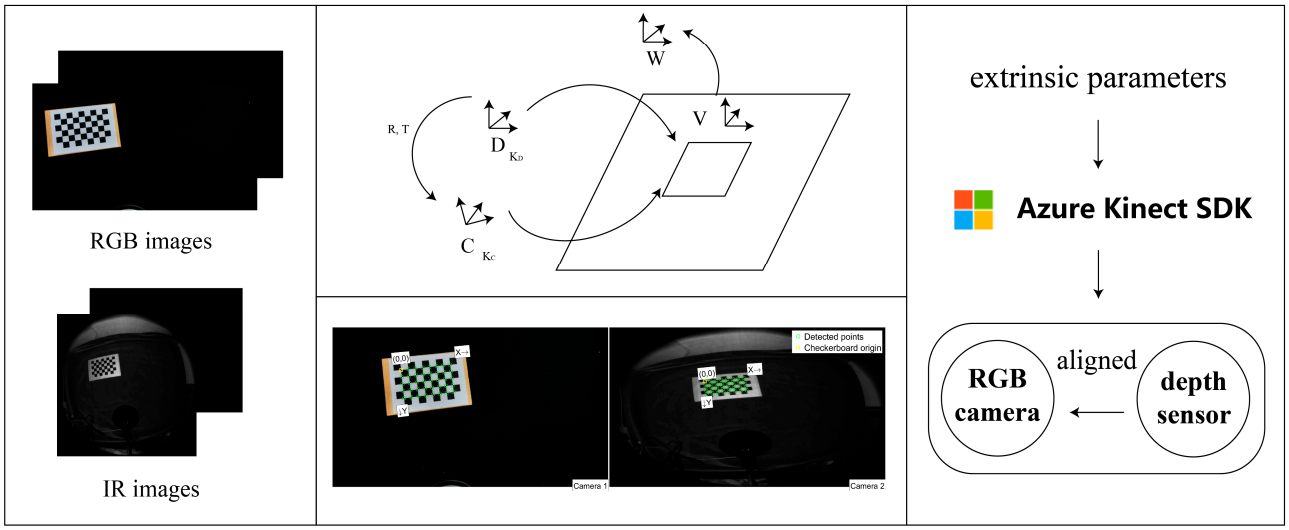


Figure 3. The recalibration strategy.

2.4. Semantic Information Guided Depth Image Inpainting

Directly repairing void pixels and flying pixels across the entire depth image is challenging and provides limited protection for leaf regions. With the growing development of computer vision, especially visual large models (LMs), nowadays, the fusion of RGB images and depth data is showing increasing potential, due to the capability of vision models to reveal semantic information in both RGB and depth data. The study proposed a semantic information-guided depth inpainting method for repairing void pixels and flying pixels in the leaf region, leveraging the multimodal coupling of RGB-D camera data to enhance depth information with organ-level semantic guidance.

To extract the semantic information of leaves in the RGB images, the study designed a leaf segmentation framework, YOLOv11-SAM2, which leverages supervised object recognition prompts and high-performance visual LMs to achieve high-precision RGB image semantic parsing with minimal manual effort. YOLO11 is a cutting-edge object detection model that builds upon the success of previous YOLO versions. To minimize manual annotations while ensuring high performance and generalization, we used the PlantDoc dataset [37] for pretraining YOLOv11 and additionally labeled 100 images of maize leaves (as mentioned in Section 2.1) with bounding boxes to fine-tune the model for improved detection in specific scenarios. This study adopted the YOLOv11-M model for training and inference, using the AdamW optimizer with an initial learning rate of 0.001, momentum set to 0.937, and a step size decay strategy. The model was trained for 100 epochs with a batch size of 8. Data augmentation techniques included horizontal flipping (probability of 0.5), random scaling (range of 0.5 to 1.5), translation (up to 10% of image size), HSV adjustments (hsv_h = 0.015, hsv_s = 0.7, and hsv_v = 0.4), and random erasing (probability of 0.4).

Segment Anything Model 2 (SAM2) is a visual LM for interactive image segmentation that returns effective masks from any given segmentation prompt [38]. After the trained YOLOv11 model conducted inference on new images, the position of each detected leaf's bounding box was subsequently fed as a prompt into SAM2 model. Finally, high-precision leaf masks were obtained from the SAM2 model. Since the accuracy of segmentation results must be guaranteed, the study selected the sam2_hiera_large model as the inference model. The whole framework of YOLOv11-SAM2 is shown in Figure 4.

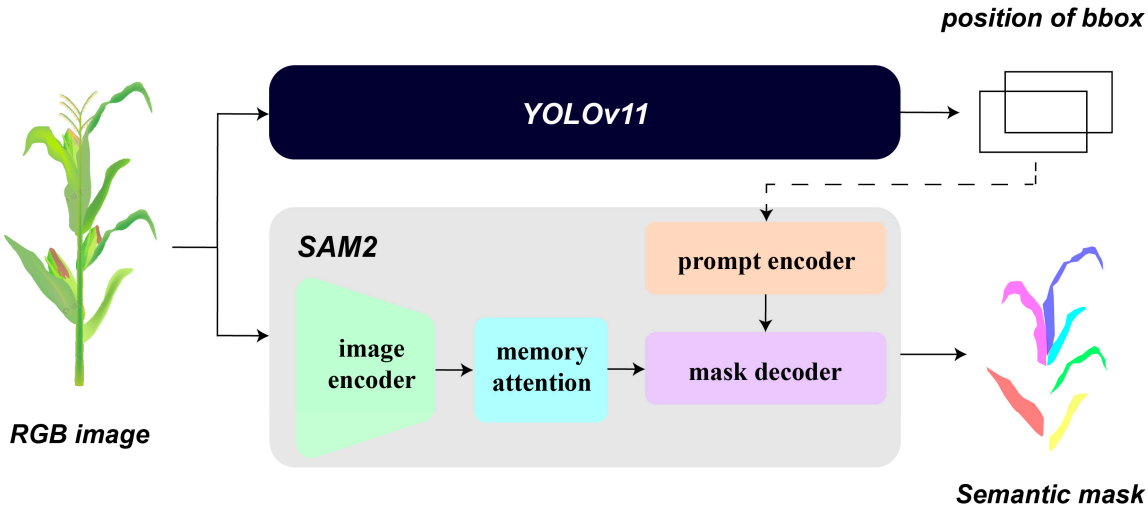


Figure 4. The framework of YOLOv11-SAM2.

After being inferred by YOLOv11-SAM2 model, the semantic results of maize leaves were applied to the aligned depth image as masks, respectively. To identify and repair void pixels and flying pixels of leaf regions in the depth image, this study proposed a depth image inpainting method by integrating innovative techniques tailored for organ-level phenotyping. Key innovations include depth inversion to prioritize high-confidence regions, topology-based edge detection to adaptively repair leaf curling and folding, and histogram filtering to dynamically eliminate anomalous pixels. The algorithm's pixel-filtering approach ensures accurate completion by leveraging spatial relationships between known and unknown pixels, while final refinement steps enhance edge continuity. These advancements significantly improve depth image quality, providing robust and reliable measurements even in complex morphological conditions. Moreover, as an unsupervised image algorithm, this method offers higher efficiency and better adaptability to diverse scenarios, as well as eliminates the need for extensive training data compared to deep-learning-based inpainting methods. The proposed method was realized via the six steps of the algorithm described below.

(1) Depth inversion. Azure Kinect DK stored depth image in 16-bit format. Objects farther from the RGB-D camera had higher values in the depth image, but with lower confidence. To prevent dilation and completion operations from allowing lower-confidence pixels to overwrite higher-confidence pixels, the depth image would first conduct depth inversion, where objects closer to the camera have higher values in the depth map. After depth inversion, the value of depth image was as follows:

$$D_{inv} = D_{\max} - D_{input} \quad (1)$$

where $D_{\max}$ was the maximum value allowed for depth image.

(2) Void pixel identification. The depth values of pixels outside the mask were set to 0, and then the void pixels inside the mask solely possess a value of $D_{\max}$. These pixels were extracted as the regions to be repaired.

(3) Skeleton-based edge pixels identification. To eliminate the flight pixels in the edge of leaf, an adaptive edge pixel identification method based on the width of leaf was adopted. Due to possible curling or folding of leaves, we first extracted the leaf skeleton using Zhang–Suen thinning algorithm from the mask. Then, the average width of the leaf was calculated using the shortest Euclidean distance from each point on the skeleton to the leaf edge. Pixels within a specific percentage of the leaf width from the leaf edge were identified as the regions to be repaired.

(4) Histogram-based anomalous pixel identification. Although the above steps removed most undesirable pixels, some anomalous flying pixels still remained undetected. Therefore, we further applied histogram filtering to eliminate these anomalous pixels. The main peak in the histogram of depth image with the largest prominence was identified, along with its left base and right base. These two values were set as the upper and lower thresholds for the depth image, allowing for the identification of anomalous pixels whose values fall outside the main peak.

(5) Depth completion based on pixel filter. For convenience, void pixels, edge pixels, and anomalous pixels are collectively referred to as unknown pixels (UPs), while other pixels within the leaf region are designated as known pixels (KPs). First, the distance matrix, which represented the distance from UPs to KPs, was calculated, prioritizing UPs that were closer to KPs. Next, the neighboring pixels of every UP within a range of R were retrieved in order. If the number of non-zero neighboring pixels exceeded threshold ∂, the depth value of the unknown pixel could be assigned as the average depth value of all non-zero neighboring pixels, as shown in Figure 5. We continued this process until all UPs were completed. This depth completion method utilized the mean approach to fully leverage the cumulative effect of neighboring pixels, ensuring that only depth values larger than those in the surrounding pixels were generated, thus providing higher confidence level in the completed depth values.

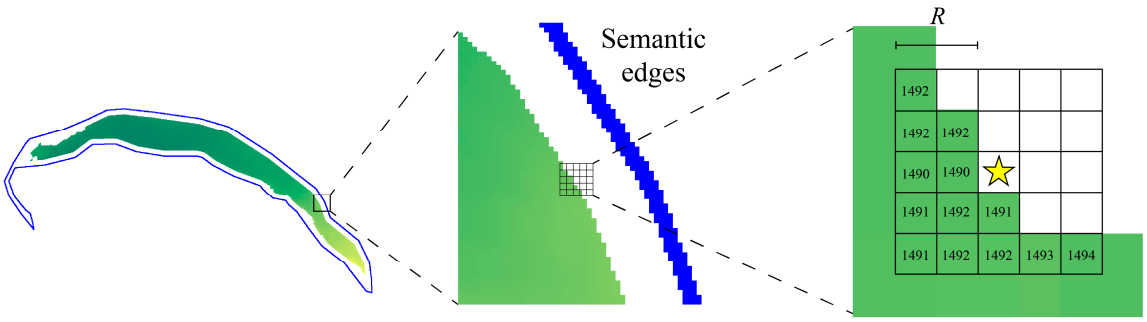


Figure 5. Pixel filter depth completion algorithm (The star in the figure represents the pixel currently being inpainted).

(6) Depth inversion. A 5 × 5 diamond-shaped kernel was used to conduct slight dilation on the depth image [39], enhancing the relationship between the edges and high-confidence pixels in the known regions. Then, the depth image was once again constrained using the semantic mask and underwent depth inversion, which was finally the completed depth image. The entire procedure for depth image inpainting is illustrated in Figure 6.

2.5. Leaf Area Measurement

To obtain leaf area measurements, the depth image needed to be mapped into three-dimensional space, followed by meshing, smoothing, and other post-processing steps. The color image and aligned depth image could be converted to point cloud by the following:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = ZK_{RGB}{}^{-1} \begin{bmatrix} v \\ u \end{bmatrix} \tag{2}$$

where X, Y, and Z are the representation of pixels mapped to coordinates in three-dimensional space. Z is the depth value of the pixel in the depth image. u and v refer to row and column of the pixel. Similarly, the RGB information of the same pixel is attached to the point (X, Y, Z).

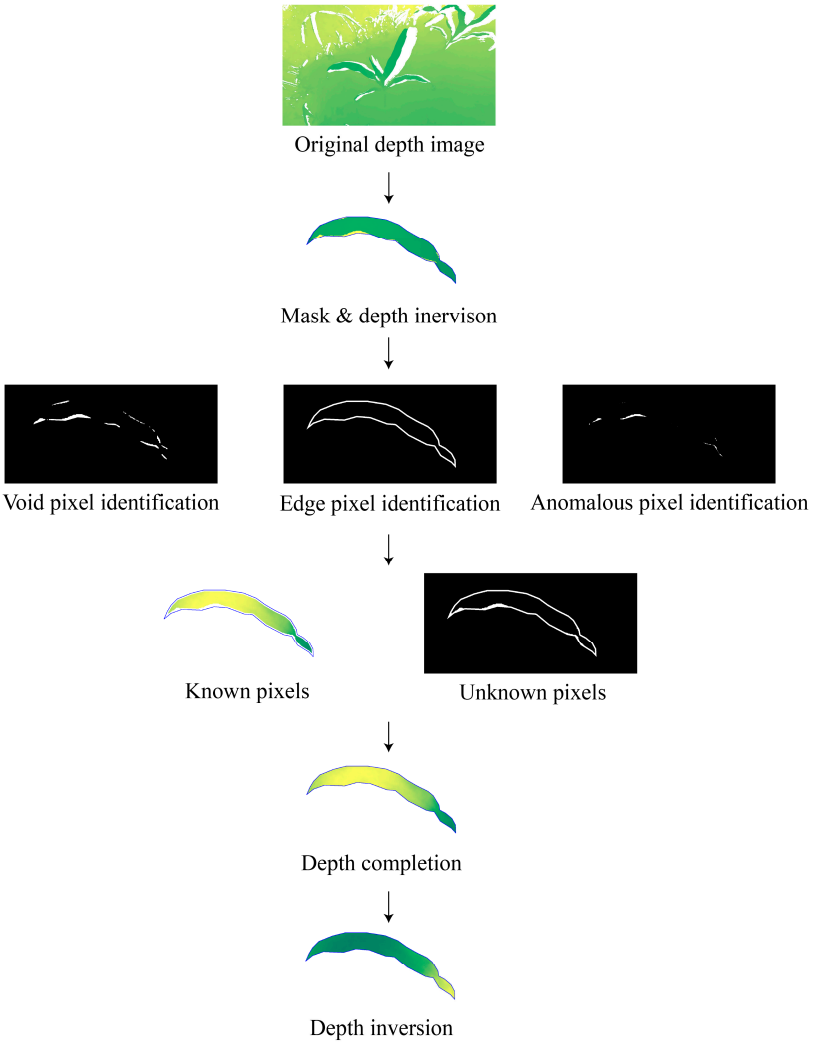


Figure 6. The procedure of depth image inpainting.

Azure Kinect DK has a resolution of 1 mm in the Z-axis. Under the sampling settings illustrated in Section 2.1, the resolutions in X-axis and Y-axis were excessively high, causing the point cloud to have a stepped shape. Since the leaf surface had a curved shape, directly measuring the area would result in significant errors. Therefore, spatial subsampling was first implemented, to ensure the point cloud was uniformly distributed with a spacing of 1 mm. Subsequently, Delaunay 2.5D triangulation was applied to the point cloud to obtain a rough meshed model of the leaf. The maximum edge length of the triangles was limited to 0.006 m. Next, the Laplacian smoothing method was applied to the mesh model, with 10 iterations and a smoothing factor set to 0.2. The leaf area could finally be measured by calculating the total area of mesh surface. The post process of 3D data for leaf area measurement is shown in Figure 7.

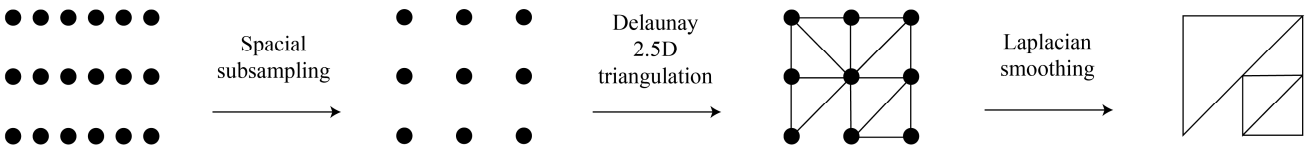


Figure 7. Example of the 3D data gridding process for leaf area measurement.

2.6. Evaluation of the Accuracy of Leaf Area Measurement

In this study, four indicators were adopted to evaluate the accuracy of leaf area measurement. Root mean square error (RMSE, Equation (3)) reflects the overall error magnitude, while mean average error (MAE, Equation (4)) represents the average absolute error. Mean absolute percentage error (MAPE, Equation (5)) highlights the relative accuracy as a percentage, and coefficient of determination (R-Square (R^2), Equation (6)) measures how well the predictions align with the actual values, with higher values indicating better performance.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{3}$$

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{4}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{6}$$

where $\hat{y}_i$ is i-th simulated value, y_i is i-th ground truth value, and $\overline{y}$ is the average of the ground truth values.

3. Results and Discussions

3.1. Recalibration Result

Derived from 20 sets of RGB and IR images, the reprojection accuracy of the external parameters of the RGB camera and depth sensor was 0.975 pixels. To facilitate the observation of alignment, we extracted the 3D point cloud of a specific maize plant generated from the RGB and aligned depth images for visualization. As shown in Figure 8c, the point cloud before the recalibration displayed noticeable misalignment, with some of the bronze color originally belonging to the ground mapped onto the leaf. In contrast, the point cloud after the recalibration (Figure 8d) possessed a consistent color, which indicated the depth image aligned closely with the RGB data. This recalibrated alignment establishes a robust foundation for utilizing RGB semantic information to guide depth map restoration in this study.

3.2. Performance of Semantic Information Guided Depth Inpainting

The yellow part of Figure 9 illustrates the semantic information extraction result of YOLOV11-SAM2. We can see that the trained YOLOv11 model accurately detected all leaves of the maize plant positioned at the center of the image view, fine-tuned and guided based on the training dataset to focus on regions with high-quality depth data. Although there were possibilities for the model to detect the leaves from other maize plants, the positions of the bounding box only from the main maize plant were fed as the prompt to the SAM2 model. The segmentation result of the SAM2 model demonstrates that, supported by the robust data engine, LMs can achieve zero-shot and high-precision semantic mask generation. The semantic masks of each leaf were applied on the depth image separately for subsequent depth inpainting, as shown in the blue part of Figure 9.

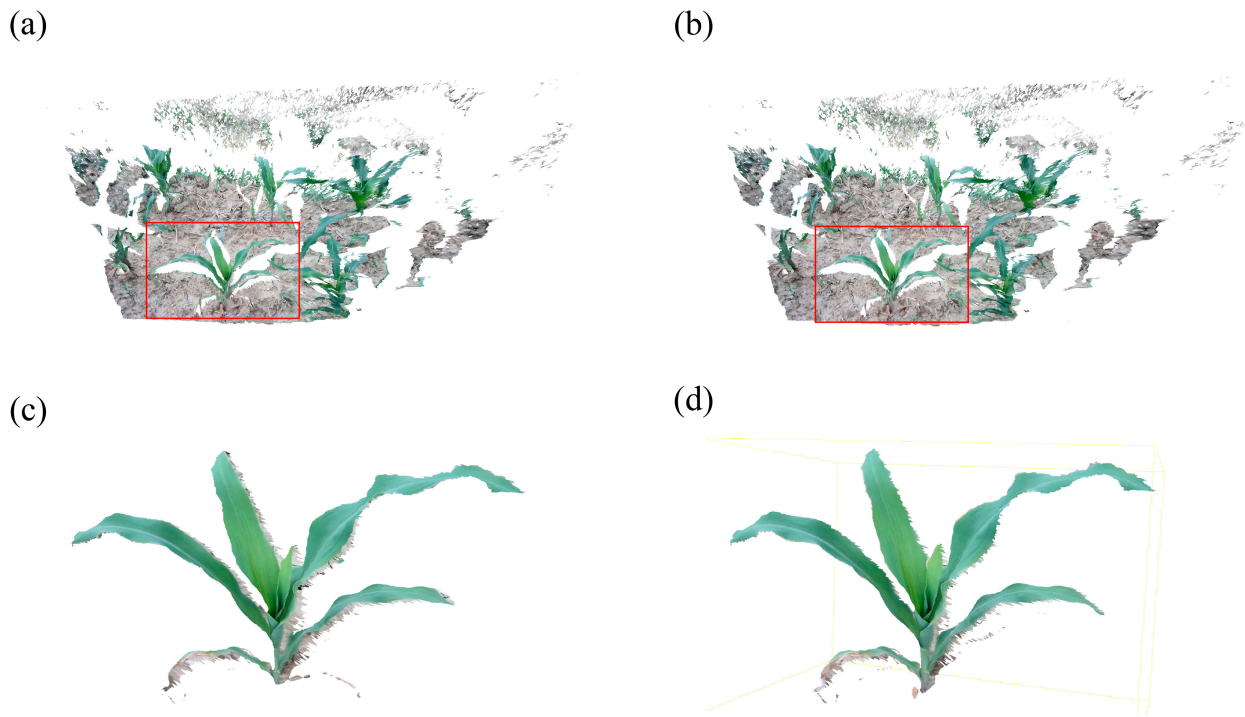


Figure 8. Point cloud of maize plant. (**a**) Total point cloud before recalibration. (**b**) Total point cloud after recalibration. (**c**) Point cloud of specific maize plant before recalibration. (**d**) Point cloud of specific maize plant after recalibration.

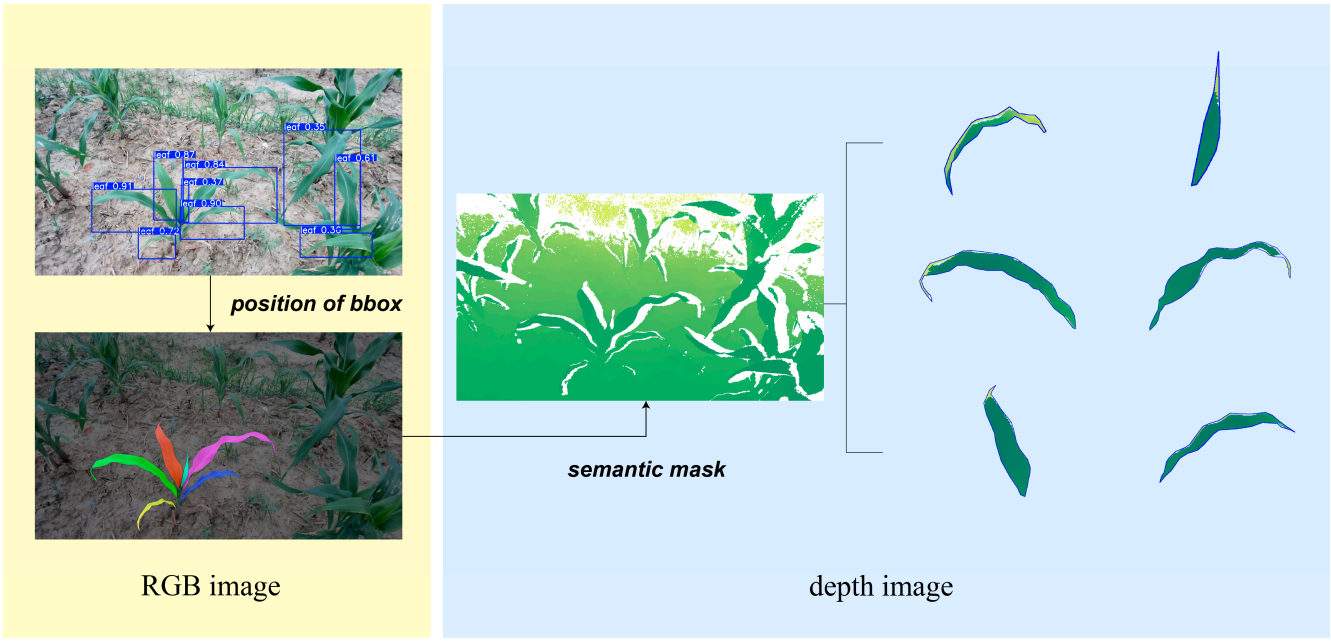


Figure 9. Semantic information extraction result of YOLOV11-SAM2 and depth image of leaf instances.

To visualize the features of the depth image, we mapped the values of the depth image to the 0–255 range and used the Summer (yellow to green) colormap to clearly distinguish the depth values. It can be observed that the UPs, including void pixels, edge pixels, and anomalous pixels, were progressively identified by the pixel identification method designed in this study.

Void pixel identification can extract most of the unsensed leaf regions from the original depth map, maximizing the restoration of leaf area. As illustrated in Figure 10a, the blue

line represents the semantic mask contour of the leaf, with the white area inside this contour highlighting the presence of void pixels. Figure 10b demonstrates that all void pixels within the leaf region have been accurately identified. After the depth inpainting, the generated point cloud in Figure 10d effectively restored the regions belonging to the leaf. Compared to the original point cloud in Figure 10c, there was a significant increase on the right side of the leaf.

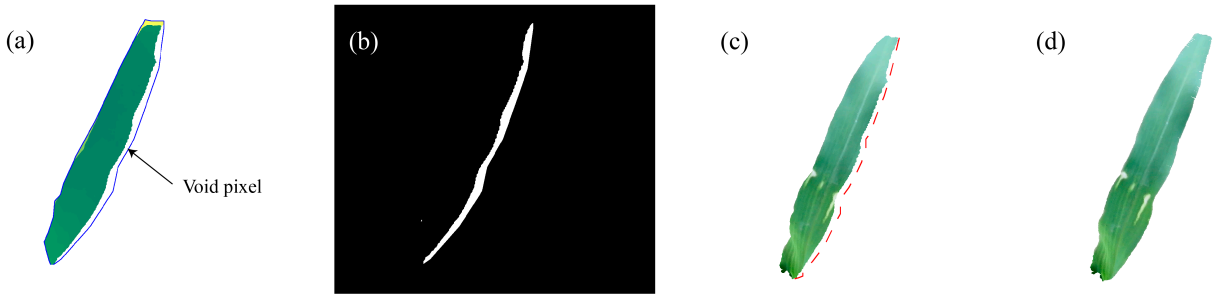


Figure 10. Void pixel identification. (**a**) Original depth image of maize leaf. (**b**) Identified void pixels. (**c**) Original point cloud of maize leaf. (**d**) Inpainted point cloud of maize leaf.

The adaptive edge pixel identification based on the leaf skeleton width can effectively remove flying pixels from the original depth image. This study tested the performance of edge pixel identification under different adaptive ratios in removing flying pixels. As shown in Figure 11, too small a ratio failed to completely remove the flying pixels, while too large a ratio removed crucial valid known pixels from the interior of the leaf. This lack of sufficient contextual information led to greater errors in depth completion. Therefore, the adaptive ratio in this study was set to 20% of the leaf skeleton width.

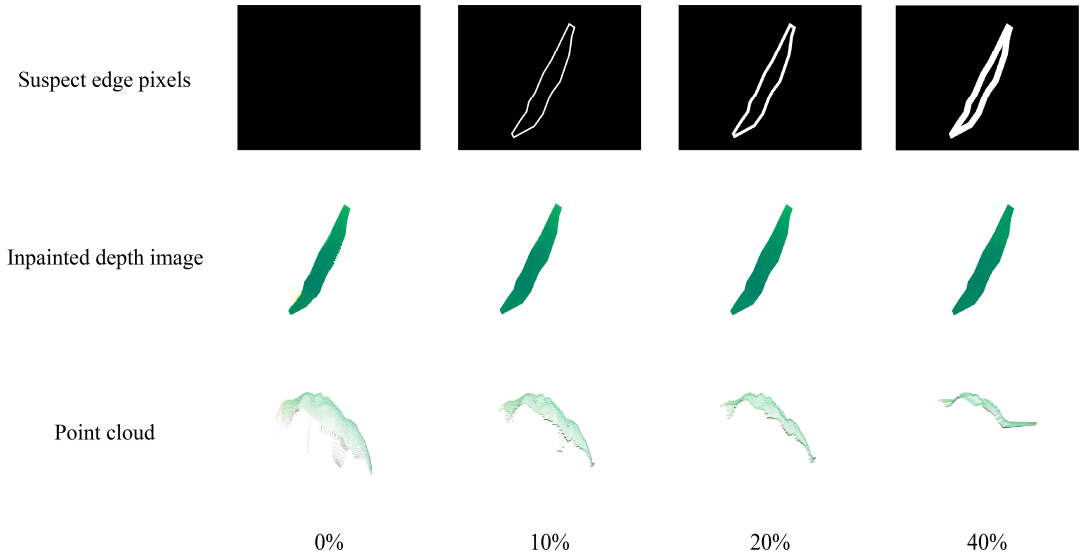


Figure 11. Visualization of inpainted depth image and point cloud under different edge identification ratios.

Histogram filtering was a crucial method for identifying anomalous pixels. Figure 12 demonstrates the effectiveness of histogram filtering for anomalous pixel identification. From Figure 12b, it can be observed that the noise points on the left side of the leaf were not detected by the above pixel identification methods, as indicated by the yellow depth values that were inconsistent with the main leaf region. Appling depth completion on these unidentified anomalous pixels could exacerbate the generation of noise points, as shown in Figure 12d. However, the histogram filtering method proposed in this study effectively

eliminated the outliers, resulting in a depth completion output with smooth gradients and less noise, as illustrated in Figure 12e. Figure 13 concretely explains that by locating the left and right bases of the histogram's main peak, anomalous pixels that crucially affect the quality of depth completion can be identified.

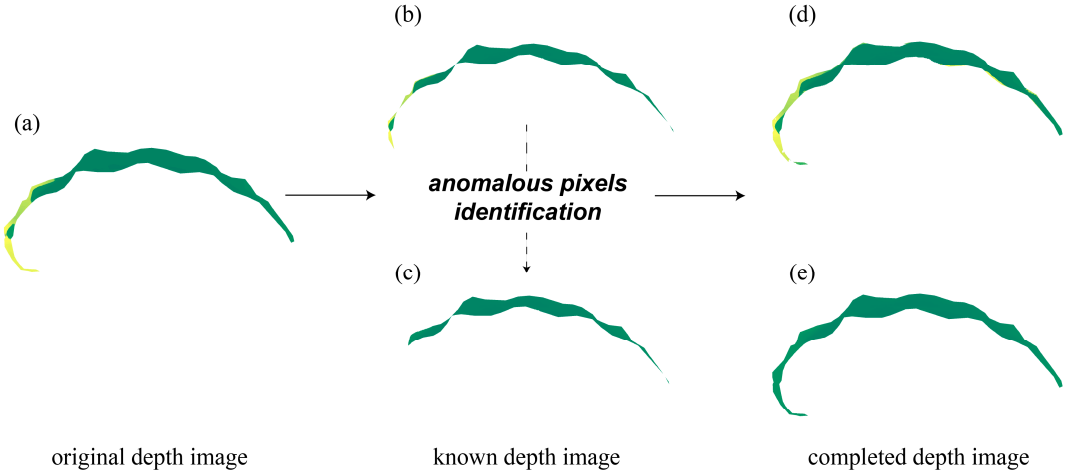


Figure 12. Visualization of anomalous pixel identification. (**a**) The original depth image of leaf. (**b**) The depth image of known pixels without anomalous pixel identification. (**c**) The depth image of known pixels with anomalous pixel identification. (**d**) The depth image after depth completion without anomalous pixel identification. (**e**) The depth image after depth completion with anomalous pixel identification.

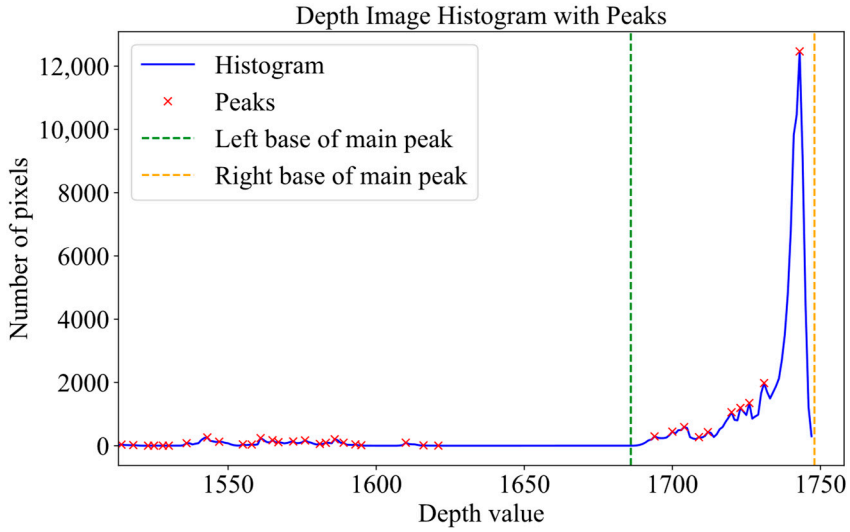


Figure 13. The histogram of depth image with peaks.

After identifying all the UPs and KPs, the pixel filter algorithm was applied for depth completion. As shown in Figure 14, the algorithm produced relatively smooth completion results, as it referenced the average values of neighboring valid pixels, thereby generating depth values with higher confidence levels and minimizing the noise introduced by high-gradient flying pixels at the edges. In terms of the search range R of the pixel filter, a smaller range (e.g., R = one) confined the algorithm to focus only on adjacent local pixels, resulting in a stripe-like completion pattern. Conversely, an excessively large range (e.g., R = 10) generated blurred depth values that were inconsistent with the original depth gradient. In this study, a radius of R = five was selected as it achieved a uniform depth that aligns with the depth gradients of the known pixels.

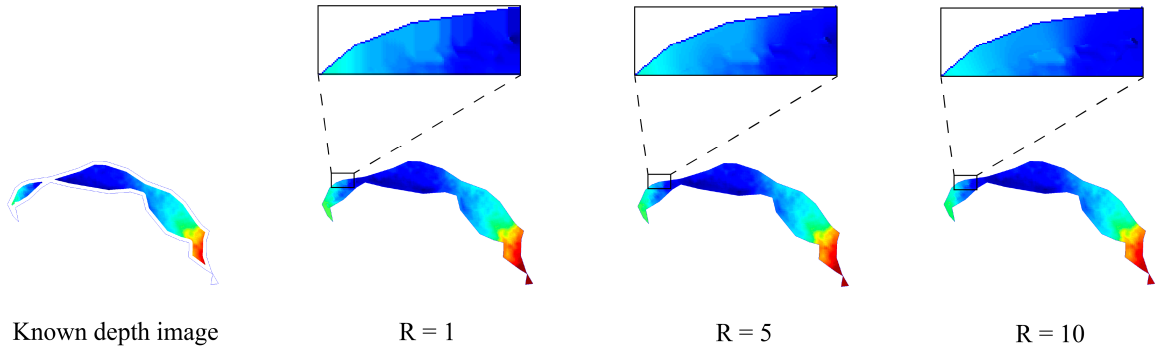


Figure 14. Performance of pixel filter using different search ranges.

3.3. Leaf Area Measurement Performance

Figure 15 illustrates the transformation process from the raw RGB-D data to the 3D mesh model that qualified for leaf area measurement. Due to the resolution limitation in the depth direction of the depth camera, the original point cloud showed a pronounced stepped distribution. After spatial subsampling, this issue was alleviated, although at the cost of some of the point cloud density. Following Delaunay triangulation, noticeable fluctuations appeared across the mesh, making the leaf model appear overly wide and thick. The final smoothed mesh, however, effectively represented both the shape and color characteristics of the leaf.

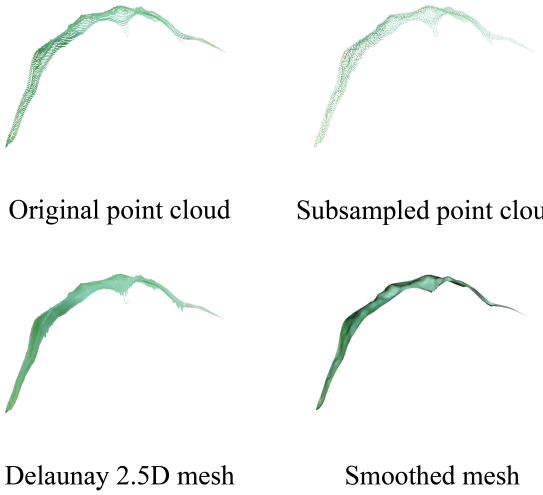


Figure 15. Visualization of 3D models for leaf area measurement.

To validate the effectiveness of the proposed method in mitigating the limitation of the Azure Kinect DK, this study also introduced the original RGB-D data for leaf area measurement as comparison. Similarly, the original data were processed by semantic masking, spatial subsampling, Delaunay 2.5D triangulation, and Laplacian smoothing.

From Table 2, it is evident that among 60 samples of leaves, the original data were significantly affected by the previously mentioned factors, resulting in larger measurement errors. In contrast, the proposed leaf area measurement method based on depth inpainting achieved an RMSE of 4.114 cm^2 and an MAE of 2.980 cm^2, while the MAPE decreased from 25.384% to 6.549%. Meanwhile, the robustness of the method was also validated by the R^2 value of 0.976. This method maintained a level of performance in field environments that was comparable to measurements taken in stable laboratory conditions, as demonstrated in several studies [23]. Figure 16 specifically illustrated the relationship between the measured leaf area and ground truth for each sample, based on our method (the red points) and the original data (the green points). Within the leaf area ranging from 10 cm^2 to 100 cm^2, the

green points were mostly scattered around both sides of the $y = x$ line, with errors primarily stemming from the overestimation caused by flying pixels and the underestimation due to the loss of leaf areas from void or anomalous depth pixels. In contrast, the red points were tightly clustered around the $y = x$ line, indicating that the method effectively mitigated the errors presented in the original data.

Table 2. The performance of leaf area measurement.

Data	RMSE/cm^2	MAE/cm^2	MAPE/%	R^2
Ours	4.114	2.980	6.549	0.976
Original	14.953	10.726	25.384	0.687

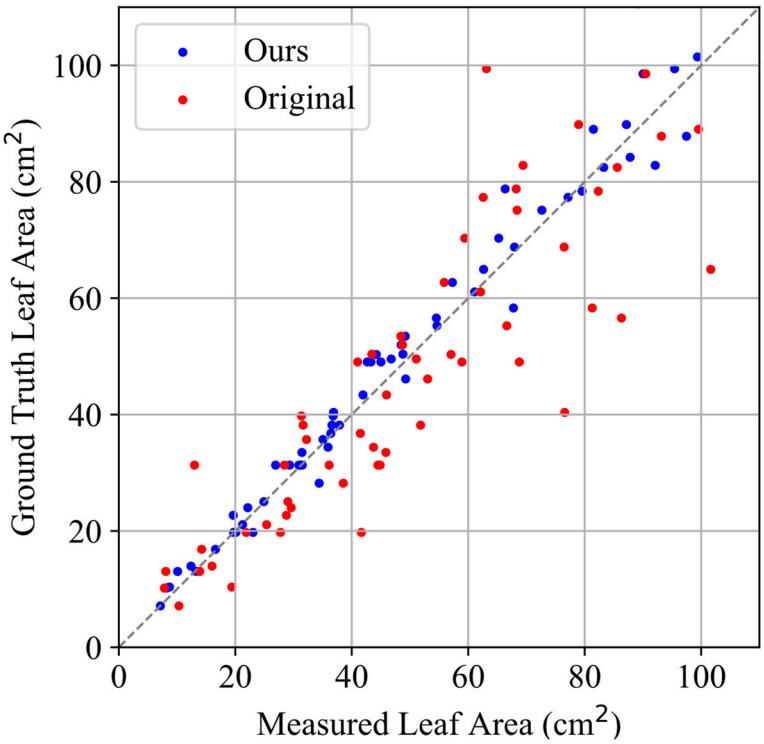


Figure 16. Scatter plots of measured and ground truth values of leaf area.

At the same time, this study validated the performance of the proposed method in measuring leaf area at different growth stages of maize plants. The collected samples from the V1–V7 growth stages were divided into early-stage samples (V1–V4) and late-stage samples (V5–V7). The early-stage samples included 25 leaf samples, while the late-stage samples included 35 leaf samples. The measurement results are shown in Table 3.

Table 3. The performance of leaf area measurement from different growth stages.

Growth Stage	Average Leaf Area/cm^2	RMSE/cm^2	MAE/cm^2	MAPE/%
V1–V4	40.907	2.274	1.812	5.824
V5–V7	60.751	5.891	3.742	6.087

As seen in Table 3, the average leaf area of the early-stage samples (V1–V4) was 40.907 cm^2, with RMSE and MAE values of 2.274 cm^2 and 1.812 cm^2, respectively, and an MAPE of 5.824%. During this stage, the leaves were smaller in size, which exhibited simpler shapes with smoother surfaces, resulting in less noise in the depth images and a

higher measurement accuracy. However, due to the smaller leaf area, the presence of flying pixels had a more noticeable impact on the relative error, causing the MAPE to approach that of the late-stage samples.

For the late-stage samples (V5–V7), the average leaf area increased to 60.751 cm^2, but the RMSE and MAE values also significantly increased to 5.891 cm^2 and 3.742 cm^2, respectively, with an MAPE of 6.087%. At this stage, the leaves exhibited an increased size, curled edges, and more intricate surface textures. These morphological features and larger measuring surfaces significantly increased the difficulty of depth measurement, leading to greater measurement errors. Despite this, the above result indicates that the proposed method exhibits good adaptability and robustness for leaf area measurement across different growth stages, even though the increased complexity of leaf morphology in the later stages is a major factor affecting the measurement accuracy.

As maize plants grow, the overall leaf area increases; however, newly emerged leaves appear simultaneously. Therefore, solely analyzing different growth stages is insufficient to validate the responsiveness of the method to leaves of varying sizes. To address this, Figure 17 presents the absolute errors (AEs) for all leaf samples, sorted by leaf area. The results show that the AE values for most samples using the proposed method were under five cm^2, demonstrating reliable performance in maize leaf area measurement across growth stages V1 to V7. However, as the leaf area increased, larger outliers in AE were frequently observed, particularly for samples with larger leaves.

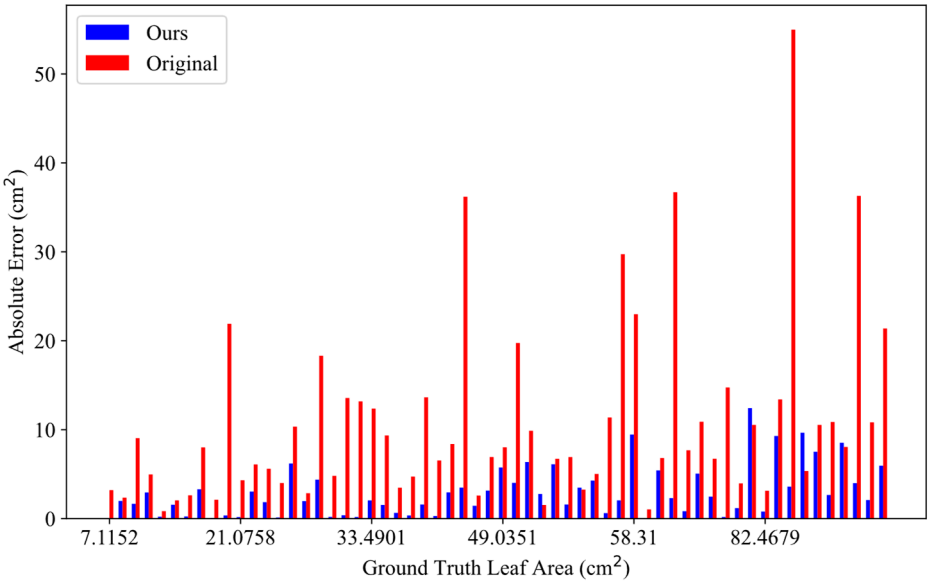


Figure 17. Absolute error plot of measured leaf area.

As previously discussed, larger measuring surfaces introduce more noise due to the inherent accuracy limitations of the RGB-D camera, contributing to systematic errors. Additionally, larger leaf edges are more prone to curling as maize plants grow, and the RGB-D camera struggles to accurately capture such complex local structures, resulting in significant errors caused by incomplete or inaccurate depth information.

To account for size differences, the absolute percentage error (APE) was used to normalize the AE. As shown in Figure 18, larger leaves exhibited relatively smaller APEs, whereas smaller leaves had higher APEs. This discrepancy arises primarily because smaller areas are more challenging for the RGB-D camera to perceive accurately, leading to the introduction of flying pixels. While the depth inpainting method mitigated some of these issues, it could not entirely remove flying pixels without compromising valid depth information, thereby increasing relative errors for smaller leaves.

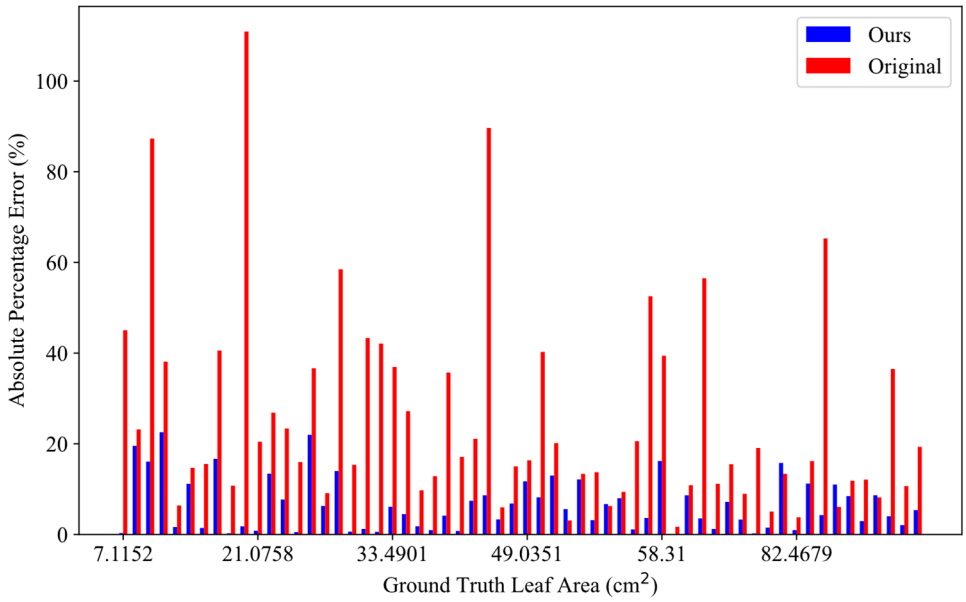


Figure 18. Absolute percentage error plot of measured leaf area.

In summary, the proposed method effectively overcomes key limitations of the Azure Kinect DK, providing a robust and adaptable performance for organ-level phenotyping, such as maize leaf area measurement, even under varying leaf sizes and morphological complexities.

4. Conclusions

Given the demand of high-throughput, organ-level 3D phenotyping for crops, particularly regarding leaf area measurement, this study parsed the common limitations of the well-known consumer-grade RGB-D camera, the Azure Kinect DK, including RGB–depth misalignment and reduced sensitivity to fine leaf structures, resulting in insufficient data quality for phenotypic analysis. To address these limitations, the study proposed a novel optimization approach for the Azure Kinect DK comprising the following components: (1) a unified recalibration protocol was developed to enhance RGB-D alignment quality, ensuring a more accurate overlay between RGB and depth data; (2) a semantic information-guided depth inpainting method was proposed based on a YOLOv11-SAM2 semantic information extraction framework, void and flying pixel identification method, and pixel filter depth inpainting algorithm; and (3) the application of maize leaf area measurement in the field was developed using the Azure Kinect DK and the optimization approach. The method finally achieved a near-laboratory level of accuracy with an MAPE of 6.549%, RMSE of 4.114 cm^2, MAE of 2.980 cm^2, and R^2 of 0.976 across 60 maize leaf samples. The application can be extended to entire plants and holds potential for measuring other crop types.

This method effectively harnesses the synergy between depth data and high-resolution RGB data provided by the Azure Kinect, employing pixel-level masks to guide depth map inpainting. By focusing operations at the image level rather than engaging in computationally intensive semantic parsing and denoising on 3D point clouds, the approach enhances both efficiency and accuracy, presenting a viable solution for high-throughput phenotypic analysis using the Azure Kinect DK. While the proposed optimization method has demonstrated a robust performance in maize leaf phenotyping, its applicability to other crops or more complex scenarios, such as dense canopies and overlapping leaves, merits further investigation. Future work could explore adaptations to address such challenges, including multi-view or multi-device data fusion for comprehensive spatial analysis

and the integration of deep learning techniques to improve precision and robustness. However, it is crucial that future developments consider operational constraints in field settings and maintain computational efficiency to ensure practical applicability in diverse agricultural contexts.

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

Funding: This research was funded by China National Key Research and Development Plan Project (2023YFD2000101) and Zhejiang Province Key Research and Development Program (2022C02056).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

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

References

1. Wang, X.; Hua, J.; Kang, M.; Wang, H.; Reffye, P. Functional–Structural Plant Model "GreenLab": A State-of-the-Art Review. *Plant Phenomics* **2024**, *6*, 0118. [CrossRef] [PubMed]
2. Zhou, L.; Zhang, H.; Bian, L.; Tian, Y.; Zhou, H. Phenotyping of Drought-Stressed Poplar Saplings Using Exemplar-Based Data Generation and Leaf-Level Structural Analysis. *Plant Phenomics* **2024**, *6*, 0205. [CrossRef]
3. Fang, H.; Baret, F.; Plummer, S.; Schaepman-Strub, G. An Overview of Global Leaf Area Index (LAI): Methods, Products, Validation, and Applications. *Rev. Geophys.* **2019**, *57*, 739–799. [CrossRef]
4. Haghshenas, A.; Emam, Y. Accelerating leaf area measurement using a volumetric approach. *Plant Methods* **2022**, *18*, 61. [CrossRef] [PubMed]
5. Ji, X.; Zhou, Z.; Gouda, M.; Zhang, W.; He, Y.; Ye, G.; Li, X. A novel labor-free method for isolating crop leaf pixels from RGB imagery: Generating labels via a topological strategy. *Comput. Electron. Agric.* **2024**, *218*, 108631. [CrossRef]
6. Yang, T.; Zhou, S.; Xu, A.; Ye, J.; Yin, J. An Approach for Plant Leaf Image Segmentation Based on YOLOV8 and the Improved DEEPLABV3+. *Plants* **2023**, *12*, 3438. [CrossRef]
7. Su, Z.; Zhou, G.; Song, L.; Lu, X.; Zhao, R.; Zhou, X. Three-Dimensional Reconstruction of Leaves Based on Laser Point Cloud Data. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 6688–6691.
8. Wen, W.; Wu, S.; Lu, X.; Liu, X.; Gu, S.; Guo, X. Accurate and semantic 3D reconstruction of maize leaves. *Comput. Electron. Agric.* **2024**, *217*, 108566. [CrossRef]
9. Ando, R.; Ozasa, Y.; Guo, W. Robust Surface Reconstruction of Plant Leaves from 3D Point Clouds. *Plant Phenomics* **2021**, *2021*, 3184185. [CrossRef]
10. Li, Y.; Wen, W.; Miao, T.; Wu, S.; Yu, Z.; Wang, X.; Guo, X.; Zhao, C. Automatic organ-level point cloud segmentation of maize shoots by integrating high-throughput data acquisition and deep learning. *Comput. Electron. Agric.* **2022**, *193*, 106702. [CrossRef]
11. Huang, T.; Bian, Y.; Niu, Z.; Taha, M.F.; He, Y.; Qiu, Z. Fast neural distance field-based three-dimensional reconstruction method for geometrical parameter extraction of walnut shell from multiview images. *Comput. Electron. Agric.* **2024**, *224*, 109189. [CrossRef]
12. Zhou, L.; Jin, S.; Wang, J.; Zhang, H.; Shi, M.; Zhou, H. 3D positioning of Camellia oleifera fruit-grabbing points for robotic harvesting. *Biosyst. Eng.* **2024**, *246*, 110–121. [CrossRef]
13. Andújar, D.; Ribeiro, A.; Fernández-Quintanilla, C.; Dorado, J. Using depth cameras to extract structural parameters to assess the growth state and yield of cauliflower crops. *Comput. Electron. Agric.* **2016**, *122*, 67–73. [CrossRef]
14. Sun, G.; Wang, X. Three-Dimensional Point Cloud Reconstruction and Morphology Measurement Method for Greenhouse Plants Based on the Kinect Sensor Self-Calibration. *Agronomy* **2019**, *9*, 596. [CrossRef]
15. Hu, Y.; Wang, L.; Xiang, L.; Wu, Q.; Jiang, H. Automatic Non-Destructive Growth Measurement of Leafy Vegetables Based on Kinect. *Sensors* **2018**, *18*, 806. [CrossRef]
16. Wu, G.; Zhu, Q.; Huang, M.; Guo, Y.; Qin, J. Automatic recognition of juicy peaches on trees based on 3D contour features and colour data. *Biosyst. Eng.* **2019**, *188*, 1–13. [CrossRef]

17. Yang, T.; Ye, J.; Zhou, S.; Xu, A.; Yin, J. 3D reconstruction method for tree seedlings based on point cloud self-registration. *Comput. Electron. Agric.* **2022**, *200*, 107210. [CrossRef]
18. Zhu, T.; Ma, X.; Guan, H.; Wu, X.; Wang, F.; Yang, C.; Jiang, Q. A calculation method of phenotypic traits based on three-dimensional reconstruction of tomato canopy. *Comput. Electron. Agric.* **2023**, *204*, 107515. [CrossRef]
19. Ma, X.; Wei, B.; Guan, H.; Cheng, Y.; Zhuo, Z. A method for calculating and simulating phenotype of soybean based on 3D reconstruction. *Eur. J. Agron.* **2024**, *154*, 127070. [CrossRef]
20. Servi, M.; Profili, A.; Furferi, R.; Volpe, Y. Comparative Evaluation of Intel RealSense D415, D435i, D455, and Microsoft Azure Kinect DK Sensors for 3D Vision Applications. *IEEE Access* **2024**, *12*, 111311–111321. [CrossRef]
21. Xie, P.; Ma, Z.; Du, R.; Yang, X.; Jiang, Y.; Cen, H. An unmanned ground vehicle phenotyping-based method to generate three-dimensional multispectral point clouds for deciphering spatial heterogeneity in plant traits. *Mol. Plant* **2024**, *17*, 1624–1638. [CrossRef] [PubMed]
22. Miranda, J.C.; Arnó, J.; Gené-Mola, J.; Lordan, J.; Asín, L.; Gregorio, E. Assessing automatic data processing algorithms for RGB-D cameras to predict fruit size and weight in apples. *Comput. Electron. Agric.* **2023**, *214*, 108302. [CrossRef]
23. Otoya, P.E.L.; Gardini, S.R.P. Real-Time Non-Invasive Leaf Area Measurement Method using Depth Images. In Proceedings of the 2020 IEEE ANDESCON, Quito, Ecuador, 13–16 October 2020; pp. 1–6.
24. Qiu, R.; Zhang, M.; He, Y. Field estimation of maize plant height at jointing stage using an RGB-D camera. *Crop J.* **2022**, *10*, 1274–1283. [CrossRef]
25. Song, P.; Li, Z.; Yang, M.; Shao, Y.; Pu, Z.; Yang, W.; Zhai, R. Dynamic detection of three-dimensional crop phenotypes based on a consumer-grade RGB-D camera. *Front. Plant Sci.* **2023**, *14*, 1097725. [CrossRef] [PubMed]
26. Boukhana, M.; Ravaglia, J.; Hétroy-Wheeler, F.; De Solan, B. Geometric models for plant leaf area estimation from 3D point clouds: A comparative study. *Graph. Vis. Comput.* **2022**, *7*, 200057. [CrossRef]
27. Chen, Q.; Huang, S.; Liu, S.; Zhong, M.; Zhang, G.; Song, L.; Zhang, X.; Zhang, J.; Wu, K.; Ye, Z.; et al. Multi-view 3D reconstruction of seedling using 2D image contour. *Biosyst. Eng.* **2024**, *243*, 130–147. [CrossRef]
28. Ma, Z.; Sun, D.; Xu, H.; Zhu, Y.; He, Y.; Cen, H. Optimization of 3D Point Clouds of Oilseed Rape Plants Based on Time-of-Flight Cameras. *Sensors* **2021**, *21*, 664. [CrossRef] [PubMed]
29. Li, Y.; Si, S.; Liu, X.; Zou, L.; Wu, W.; Liu, X.; Zhang, L. Three-dimensional reconstruction of cotton plant with internal canopy occluded structure recovery. *Comput. Electron. Agric.* **2023**, *215*, 108370. [CrossRef]
30. Chen, H.; Liu, S.; Wang, C.; Wang, C.; Gong, K.; Li, Y.; Lan, Y. Point Cloud Completion of Plant Leaves under Occlusion Conditions Based on Deep Learning. *Plant Phenomics* **2023**, *5*, 0117. [CrossRef]
31. Yang, R.S.; Chan, Y.H.; Gong, R.; Nguyen, M.; Strozzi, A.G.; Delmas, P.; Gimel'farb, G.; Ababou, R. Multi-Kinect scene reconstruction: Calibration and depth inconsistencies. In Proceedings of the 2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013), Wellington, New Zealand, 27–29 November 2013; pp. 47–52.
32. Wang, Z.; Song, X.; Wang, S.; Xiao, J.; Zhong, R.; Hu, R. Filling Kinect depth holes via position-guided matrix completion. *Neurocomputing* **2016**, *215*, 48–52. [CrossRef]
33. Paredes, A.L.; Song, Q.; Conde, M.H. Performance Evaluation of State-of-the-Art High-Resolution Time-of-Flight Cameras. *IEEE Sens. J.* **2023**, *23*, 13711–13727. [CrossRef]
34. Tölgyessy, M.; Dekan, M.; Chovanec, L'.; Hubinský, P. Evaluation of the Azure Kinect and Its Comparison to Kinect V1 and Kinect V2. *Sensors* **2021**, *21*, 413. [CrossRef] [PubMed]
35. Lachat, E.; Macher, H.; Landes, T.; Grussenmeyer, P. Assessment and Calibration of a RGB-D Camera (Kinect v2 Sensor) Towards a Potential Use for Close-Range 3D Modeling. *Remote Sens.* **2015**, *7*, 13070–13097. [CrossRef]
36. Wei, F.; Xu, G.; Wu, Q.; Kuang, J.; Tian, P.; Qin, P.; Li, Z. Azure Kinect Calibration and Parameter Recommendation in Different Scenarios. *IEEE Sens. J.* **2022**, *22*, 9733–9742. [CrossRef]
37. Singh, D.; Jain, N.; Jain, P.; Kayal, P.; Kumawat, S.; Batra, N. PlantDoc: A Dataset for Visual Plant Disease Detection. In Proceedings of the CoDS COMAD 2020, Hyderabad, India, 5–7 January 2020; pp. 249–253.
38. Ravi, N.; Gabeur, V.; Hu, Y.-T.; Hu, R.; Ryali, C.; Ma, T.; Khedr, H.; Rädle, R.; Rolland, C.; Gustafson, L.; et al. SAM 2: Segment Anything in Images and Videos. *arXiv* **2024**, arXiv:2408.00714.
39. Ku, J.; Harakeh, A.; Waslander, S.L. In Defense of Classical Image Processing: Fast Depth Completion on the CPU. In Proceedings of the 2018 15th Conference on Computer and Robot Vision (CRV), Toronto, ON, Canada, 8–10 May 2018; pp. 16–22.

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.