Novel 3D Imaging Systems for High-Throughput Phenotyping of Plants

Tian Gao [1], Feiyu Zhu [1], Puneet Paul [2], Jaspreet Sandhu [2], Henry Akrofi Doku [2], Jianxin Sun [1], Yu Pan [1], Paul Staswick [2], Harkamal Walia [2] and Hongfeng Yu [1,*]

[1] Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588, USA; tgao@huskers.unl.edu (T.G.); feiyu.zhu@huskers.unl.edu (F.Z.); jianxin.sun@huskers.unl.edu (J.S.); yu@huskers.unl.edu (Y.P.)

[2] Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583, USA; puneetpaul@unl.edu (P.P.); jsandhu@huskers.unl.edu (J.S.); henry.doku@huskers.unl.edu (H.A.D.); pstaswick1@unl.edu (P.S.); hwalia2@unl.edu (H.W.)

[*] Correspondence: hfyu@unl.edu; Tel.: +1-402-472-5013

Abstract: The use of 3D plant models for high-throughput phenotyping is increasingly becoming a preferred method for many plant science researchers. Numerous camera-based imaging systems and reconstruction algorithms have been developed for the 3D reconstruction of plants. However, it is still challenging to build an imaging system with high-quality results at a low cost. Useful comparative information for existing imaging systems and their improvements is also limited, making it challenging for researchers to make data-based selections. The objective of this study is to explore the possible solutions to address these issues. We introduce two novel systems for plants of various sizes, as well as a pipeline to generate high-quality 3D point clouds and meshes. The higher accuracy and efficiency of the proposed systems make it a potentially valuable tool for enhancing high-throughput phenotyping by integrating 3D traits for increased resolution and measuring traits that are not amenable to 2D imaging approaches. The study shows that the phenotype traits derived from the 3D models are highly correlated with manually measured phenotypic traits ($R^2 > 0.91$). Moreover, we present a systematic analysis of different settings of the imaging systems and a comparison with the traditional system, which provide recommendations for plant scientists to improve the accuracy of 3D construction. In summary, our proposed imaging systems are suggested for 3D reconstruction of plants. Moreover, the analysis results of the different settings in this paper can be used for designing new customized imaging systems and improving their accuracy.

Keywords: 3D reconstruction; point cloud; imaging system; high-throughput phenotyping

1. Introduction

High-throughput phenotyping is a critical component of plant science research aimed at improving crop performance for meeting the food, fiber, and fuel needs of society. Accurate and rapid quantification of plant phenotypes can enable researchers to bridge the genotype-to-phenotype gap, especially for traits associated with stress tolerance [1]. High-throughput phenotyping has the potential to accelerate the development of high-yielding, stress-tolerant crops [2].

It is challenging to develop cost-effective high-throughput phenotyping systems. One of the popular existing solutions is image-based methods. Compared to manual phenotyping, which is laborious, time-consuming, and usually destructive [2], image-based methods are desired for their efficiency, non-destructive aspect, and the capability of large-scale measurements. For example, Zhou et al. [3] presented a semi-automated phenotyping pipeline named Toolkit for Inflorescence Measurement (TIM) to extract traits from images of sorghum. Gage et al. [4] developed a Tassel image-based phenotyping system (TIPS) for tassel imaging in the field. Although image-based methods have successful applications

in phenotyping, there are still numerous limitations that could become the barrier for wider adoption by researchers. As images are projections of 3D objects onto 2D planes, image-based methods cannot present an accurate structural description of 3D objects due to the occlusion and inevitable loss of depth information. As a result, extra efforts are needed to estimate the spatial information of plants in 3D space [5]. Moreover, captured images are related to view-angles, and, thus, the traits of the same plant might be different if the spatial relationship between the camera and the plant changes. This instability leads to inaccuracies during phenotyping and makes it difficult for researchers to draw reliable conclusions on the genotype-to-phenotype linkages.

In order to overcome some of the drawbacks of image-based methods, plant scientists are exploring available 3D approaches for improving phenotyping. Compared to images, 3D models of the plants (usually represented as point clouds or meshes) include the depth information intrinsically. Therefore, 3D models have shown the promising capacity to describe the complete spatial information of the plants and, thereby, avoid the issues of view-angle dependent traits. Moreover, similar to the image-based methods, 3D methods can also be non-destructive and scalable for phenomics experiments. In general, there are two main types of methods to reconstruct a plant in 3D space. The first is active methods, in which various sensors transmit and receive signals actively to capture the depth information. In these methods, plants are scanned from multiple view-angles to generate raw angle-specific point clouds. Then, the raw point clouds are registered and merged to construct the final point clouds. The advantage of this type of method is its easy access to 3D point clouds. For example, Thapa et al. [6], and Zhu et al. [7] proposed an instrument based on light detection and ranging (LiDAR) to capture the point clouds of maize and sorghum. The second type entails passive methods, which only involve 2D images captured by regular cameras. With the images from various view-angles, the depth information is calculated, and the 3D shapes of plants are reconstructed using various algorithms. One of the most favored algorithms is structure-from-motion (SfM), in which the positions of points of the plants in 3D space are calculated by constructing a 3D scene using paired images [8]. Another algorithm, multi-view environment (MVE), combines SfM and multi-view stereo (MVS) algorithms together and reconstructs a point cloud and 3D meshes [9]. MVE has been used to develop a pipeline for 3D reconstruction and phenotyping to study growth dynamics of rice inflorescences [10,11].

Although 3D phenotyping is typically more accurate than 2D image-based approaches, current 3D methods are limited due to several challenges. For example, McCormick et al. [12] proposed a pipeline to identify shoot architecture based on depth images captured by Microsoft Kinect. However, the average point spacing of Microsoft Kinect is 5 mm, while the diameter of an awn on a spike is less than 1 mm [13]. As a result, the awns can be easily considered as noise and erroneously removed in the reconstruction process. Cao et al. [14] developed a 3D imaging system with a stepper-motor-controlled frame and a regular camera for 3D reconstruction of soybean using SfM. This low-cost imaging system cannot be directly applied to plants with complex structures due to occlusion problems, since only 20 images are captured for each plant. Insufficient images and the occlusion caused by the proximal organs or leaves of the complex plant structure will lead to an incomplete 3D model and, hence, inaccurate phenotypes. He et al. [15] built an imaging system with a turntable to phenotype strawberries from local supermarkets. Here, the strawberries are placed on the center of a turntable and rotated at a certain speed while cameras capture images at a fixed position. Although such an imaging system has proved its potential in reconstructing 3D models of rigid objects such as strawberries, it is not suitable for plants with non-rigid tissues such as leaves. Since the leaves will vibrate due to the motion of the plants, a high level of noise is inevitable when generating 3D models, especially at the leaf tips. Chaudhury et al. [16] proposed an imaging system with a robot arm holding a range scanner controlled by software. Nguyen et al. [17] built a 3D reconstruction system with a mechanical arm holding 10 cameras controlled by a software application. Wu et al. [18] generated their point clouds using multiple depth cameras and de-noising algorithms.

Although they eliminate the vibration problem, these imaging systems are either too expensive or not easily accessible as a sophisticated mechanical set-up is required. More importantly, each of these imaging systems was designed for a specific plant, which may not be optimal for a different plant species. To the best of our knowledge, most existing work mainly focused on designing and implementing an imaging system for a specific type of plant and evaluating the quality of measurements obtained by the imaging system but lacked comparisons between different imaging systems.

We have developed two new controlled environments imaging systems for plants (ISP), and proposed an end-to-end pipeline to generate de-noised point clouds. Our previous work has shown the potential of our imaging systems to capture dynamics of developing plants [10,11]. Our imaging systems designed for high-throughput phenotyping are adaptable to various plant sizes with high accuracy and flexibility at a low cost. In this paper, we conduct a systemic analysis on the settings of our systems, and present a comparison study with the traditional turntable-based image system. With the hypothesis that our imaging systems are accurate enough to estimate the phenotypic traits, we extend our work and design correlation analysis on manually measured data and estimated data from 3D models. By constructing the 3D model of the same plant with various settings, we discuss how different parameters, such as the checkerboards in the imaging system and the number of images, affect the performance and the results from the presented systems. Further, by comparing results with a standard turntable-based imaging system, we provide insights on the performance and present evidence for increased accuracy from our imaging systems.

2. Materials and Methods

2.1. Setting and Materials

We performed multiple imaging settings and materials to optimize the experimental set-up by identifying the key factors affecting the quality of the final results. Inspired by the colorful Rubik's cube used in the existing imaging system [19], we utilized specially designed color checkerboards to improve the quality of the reconstructed 3D models for this optimization. Black backdrops and black paint were used for the imaging systems described in this work.

2.1.1. Camera Setting

Two digital color cameras (Sony α 6500, Sony Inc., Tokyo, Japan) were used to capture multi-view images. With a camera built-in application called "Time-lapse" [20], images were captured sequentially at the rate of up to one image per second. The total number of images can also be adjusted, with each camera capable of capturing up to 60 images per minute with a resolution of 6000 pixels $\times$ 4000 pixels per image.

2.1.2. Color Checkerboards

The specially designed color checkerboards consisted of 20×20 squares with randomly distributed colors in RGB color space. The size of each square was 1 cm^2. They were placed around the target object to provide extra image features. Image features were pieces of local information in an image, and they were used to find correspondences in paired images and help in recovering the camera parameters in a 3D scene [9]. Because of the size, the relatively uniform color, and the irregular texture, the number of image features detected in the region of plants themselves was relatively limited. As a result, the parameters of cameras, such as position and orientation, cannot be correctly recovered, which may result in apparent errors in the generated point clouds. On the other hand, due to the randomly generated color and regular square shape, the image features (usually located at corners or edges of each square) can be easily detected by feature detection algorithms. These image features provided additional correspondences and led to more accurate and stable point clouds.

2.1.3. Black Backdrops and Black Paint

Black backdrops and black paint were aimed at blocking the objects that were not of interest. If captured in images, these objects will also be constructed in the 3D scene and, thus, slow down the 3D reconstruction process. Moreover, the backdrops and paint also facilitated the selection of thresholds in the image preprocessing step.

2.2. Imaging Systems

We built two novel imaging systems according to the size of the imaged objects. For comparison, we also built a typical turntable-based imaging system.

2.2.1. Our Imaging System for Whole Plants

The first imaging system we developed was for reconstructing maize, and it can also be applied to any plant up to 2 m in height. As shown in Figure 1, we used a double-ring Lazy Susan turntable ring apparatus in the center of the system. The maize plant grown in a pot was placed detachedly in the middle of the ring apparatus. The ring apparatus in the center had two layers. The inner layer was fixed on the floor, while the outer layer was attached to the end of a flat wooden board and rotated freely. On the other side of the wooden board, a robotic car was attached to provide the power for rotation. On the wooden board, there were two tripods holding the cameras at different adjustable heights. The view-angles of the two cameras toward the plant were 45° and 30° with respect to the horizontal direction. The cameras were set with ISO value at 1250, shutter speed at 1/50 s, and aperture value at f/22. The system included black backdrops and checkerboards mentioned above. The black backdrops were around the apparatus, and the checkerboards were placed on the ground around the target plant. Figure 2a shows the photograph of the imaging system.
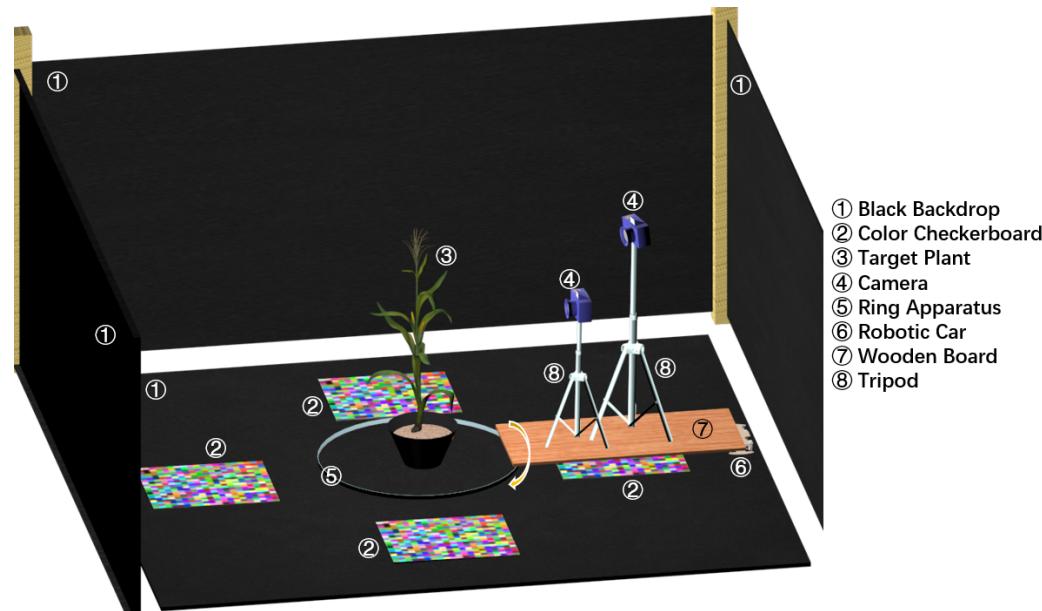

Figure 1. Our first imaging system for whole plants.

2.2.2. Our Imaging System for Targeted Plant Organs

The first imaging system worked perfectly for the whole plant on a large scale. However, it was not compatible if the area of interest was a specific plant tissue or organ such as a rice panicle/inflorescence as the target tissue of interest was relatively small and usually occluded by leaves. As a result, it was difficult to generate 3D models of acceptable quality for panicles because the image features on panicles cannot be detected. To address the occlusion issue, we developed a second imaging system specially designed for small tissues of plants. As illustrated in Figure 3, we built a wooden table with a circular board in a

customized wooden chamber to host the imaging system. Similar to the first imaging system, a double-ring Lazy Susan turntable ring apparatus was installed on the wooden board. The inner layer of the ring apparatus was fixed while the outer layer was attached to a small wooden platform holding two mini tripods, the cameras, and a LED light (ESDDI PLV-380, 15 Watt, 5000 LM, 5600 K). The cameras generated images with ISO value at 1600, shutter speed at 1/30 s, and aperture value at f/22. An electric motor system was connected to the outer layer of the apparatus with a timing belt to provide power for rotation. The electric motor system consisted of two parts: (i) a high torque motor powered by a DC power supply; (ii) an idler pulley to move the timing belt. When the power was on, the wooden platform rotated along with the ring apparatus as the belt moved. The top surface of the circular wooden board and the interior of the chamber were painted black. Color checkerboards were attached to the top surface of the circular wooden board and the chamber interior. When imaging panicles, we placed the plant under the table and passed the target panicle through the hole in the center of the board. A desktop motorized adjustable computer stand was adapted for height adjustment to keep the panicles at a similar height. The photograph of the imaging system is demonstrated in Figure 2b.
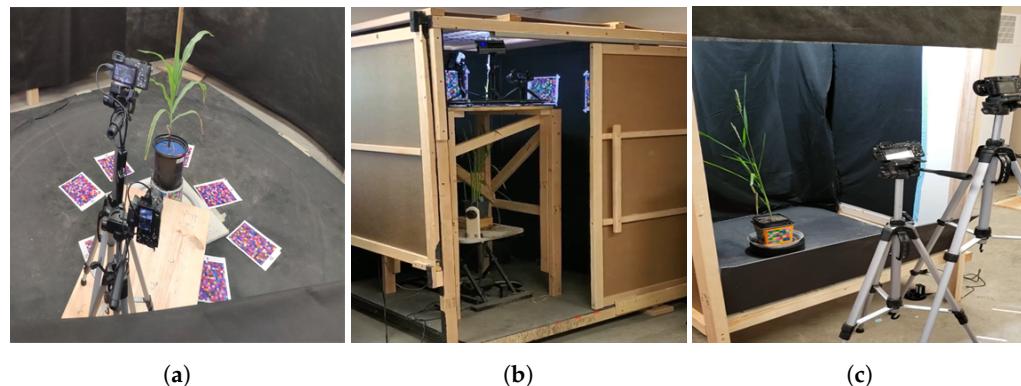

(**a**)   (**b**)   (**c**)

Figure 2. The photograph of the three imaging systems: (**a**) our imaging system for whole plants; (**b**) our imaging system for small tissues; and (**c**) a conventional turntable-based imaging system.

① Wooden Chamber
② Wooden Table
③ Target Plant
④ Cameras & LED Light
⑤ Adjustable Stand
⑥ Mini Tripod
⑦ Ring Apparatus
⑧ Timing Belt
⑨ Pulley & Motor
⑩ Color Checkerboard

Figure 3. Our second imaging system for panicles.

2.2.3. Turntable-Based Imaging System for Whole Plants

We also built a typical turntable-based imaging system for comparison. As demonstrated in Figure 4, a turntable with a plant was placed in a wooden chamber. Cameras

installed on tripods were placed outside the chamber facing the plants with the same view-angle as in the first imaging system. Similar to the second imaging system, the interior of the chamber was painted black. Instead of being attached to the chamber, the color checkerboards were cut and attached to the pot because the construction of the scene requires the static spatial relationship between the checkerboard and the plant. When the turntable was turned on, the plant began to rotate, and the rotating speed was constant in the imaging process. The photograph of this imaging system is shown in Figure 2c.
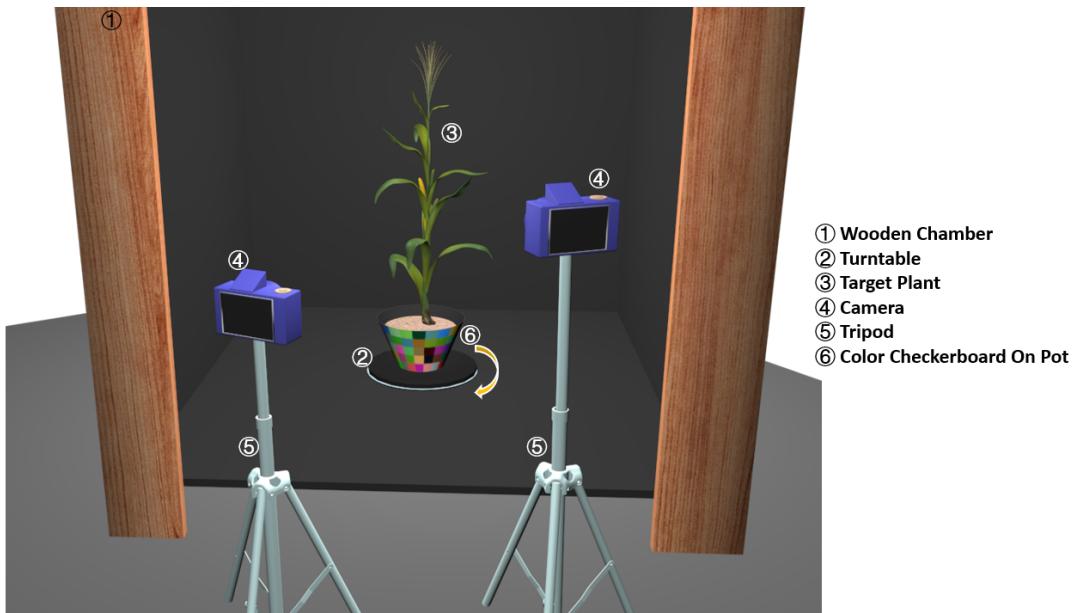

Figure 4. Conventional turntable-based imaging system.

2.3. Point Cloud Reconstruction

We reconstructed 3D point clouds from 2D plant images generated by the above-described imaging systems. Our reconstruction pipeline consisted of three main steps: image pre-processing, 3D point cloud reconstruction, and point cloud post-processing. Our pipeline was applicable for both whole plants and targeted organs, and we detail its use for rice panicle reconstruction as an example.

2.3.1. Image Preprocessing

Before 3D reconstruction, the images needed to be preprocessed to remove the background. In this work, we conducted preprocessing by employing filtering and thresholding in the color space. The goal of filtering was to remove the pixels in the background to speed up the process of 3D reconstruction since fewer pixels were utilized. However, since the distribution of the pixels in the raw images was relatively uniform in the red, green, and blue (RGB) color space, it was challenging to select an effective color threshold. Therefore, we transformed the original images into the hue, saturation, and value (HSV) color space and filtered out pixels using thresholding on the HSV channels. In this work, pixels were removed if values of their hue, saturation, and value channels were not in the ranges of 0–1, 0–1, and 0.136–1, respectively. After color thresholding, a few pixels in the background still remained and were sparsely distributed. These pixels were considered in the 3D reconstruction pipeline as outliers and ignored.

2.3.2. 3D Reconstruction

To reconstruct an accurate dense point cloud from the images, we implemented MVE [9] in this work. Figure 5 shows the pipeline of MVE. The input of MVE was the preprocessed images captured from various view angles shown in Figure 5a. First, SIFT [21] and SURF [22] were performed on these images to detect image features. An

example of images with detected image features was illustrated in Figure 5b, in which features were marked as red points (detected by SIFT) and green points (detected by SURF). Then, the parameters of the cameras, such as orientation and position, were recovered by matching the corresponding image features. As shown in Figure 5c, a 3D scene was built with these recovered parameters, including a sparse point cloud and all the camera positions. The number of camera positions in the scene matched the number of the input images. After that, a depth map (Figure 5d) was generated for each image by calculating the depth information of each pixel. Then, a dense point cloud (Figure 5e) was produced by merging all these depth maps. Finally, FSSR [23] was employed on the dense point cloud to generate the de-noised point cloud as well as the mesh (Figure 5f).
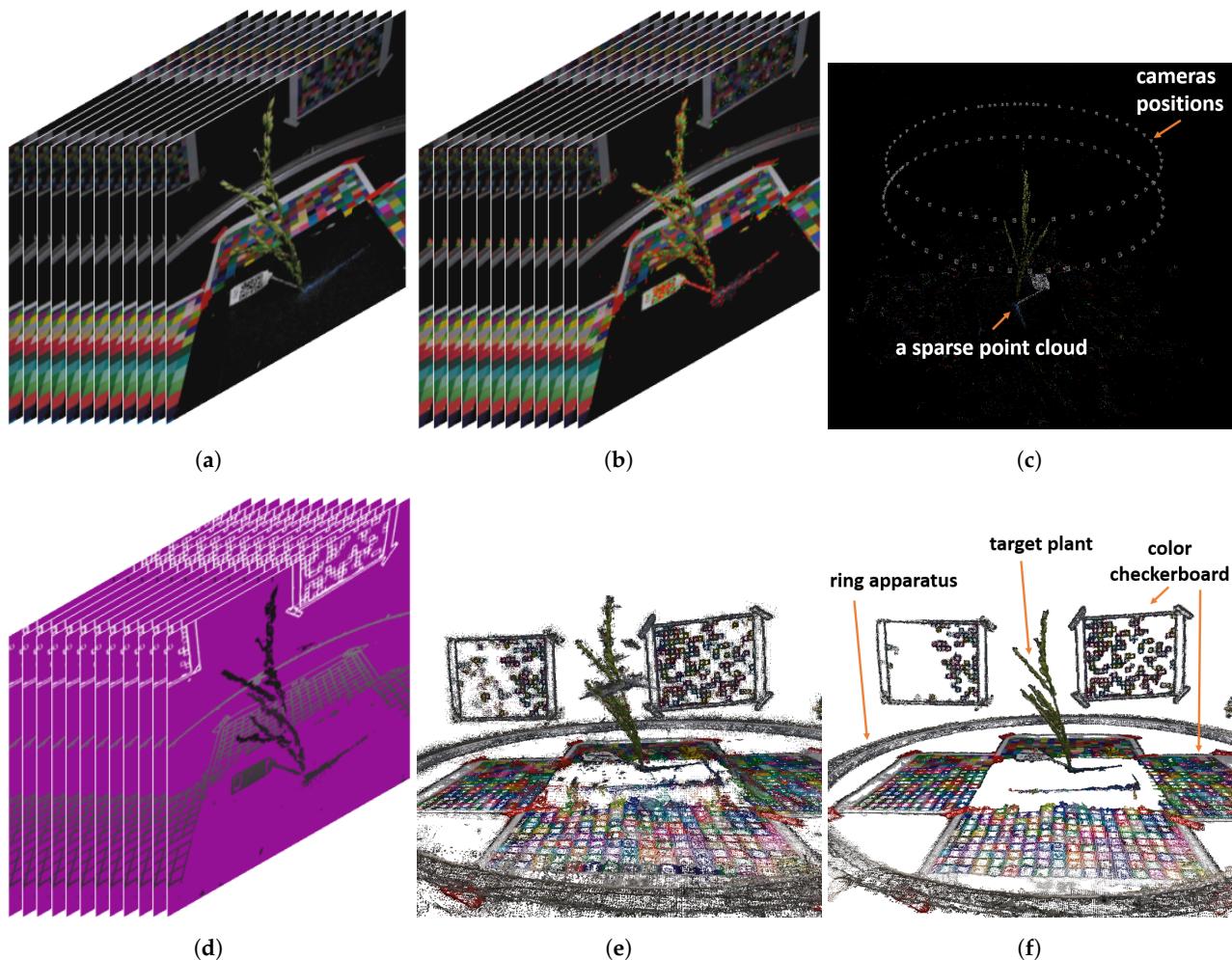

(a)   (b)   (c)

(d)   (e)   (f)

Figure 5. The pipeline of MVE: (**a**) input images captured from a set of view angles; (**b**) images with detected image features, where red and green points indicate image features detected by SIFT and SURF, respectively; (**c**) reconstructed 3D scene including camera positions and a sparse point cloud; (**d**) depth maps; (**e**) a dense 3D point cloud; and (**f**) a de-noised point cloud.

2.3.3. Point Cloud Post-processing

Since the 3D model of a target plant was the only object of interest, the other parts, including the checkerboards in the point clouds, needed to be removed. As demonstrated in Figure 6, a two-step filtering was implemented to generate the final point cloud. Panicles were utilized to illustrate the filtering process in this section. The first step was to segment the plant from the background. We performed 3D clustering on the de-noised point cloud generated by MVE (Figure 6a). The clustering algorithm was G-DBSCAN [24], which was integrated with MATLAB function "pcsegdist". Euclidean distance was used as the distance

metric for clustering. Then, we set criteria to identify the points belonging to the target plant, and denoted these points as *target points*. Intuitively, these points were generally green. Thus, we implemented the criteria by setting a threshold on visible atmospherically resistant index (VARI) [25]. VARI is one of the most popular vegetation indices for remote sensing leaf chlorophyll content. VARI has been widely used in agricultural monitoring and vegetation detection [26–28]. Compared with green band information, these vegetation indices can reduce the variations due to extraneous factors such as ambient lights [29]. The formula of VARI was illustrated in Equation (1),

$$VARI = \frac{G - R}{G + R - B} \tag{1}$$

where R, G, and B represent the values in red, green, and blue channels in the RGB color space of a point, respectively. The existing studies utilized various VARI values corresponding to a wide range of green class [30]. In this work, the panicle cluster was the main green object in our controlled environment, and, thus, we only need to estimate a VARI range to distinguish the color of the panicle cluster from the background. According to our empirical study, we set the threshold of VARI to 0.1 and marked points with VARI greater than this threshold as target points. Then, we counted the number of total points and target points within each cluster. With this threshold, we can successfully detect the panicle cluster as it had the highest percentage of the target points. The rest of the clusters, which belonged to the background, checkerboards, and the ring apparatus, were removed. However, since labels with barcode identifier, which were widely used in high-throughput phenotyping, were attached to the panicle as shown in Figure 6b, they cannot be filtered out in the first step. The second step was designed to remove these labels. We first identified all the target points using VARI again and removed them. Since the points that belonged to the labels were not target points, these points remained. After that, we fit these points to a plane as the labels were placed on the table and flat. Then, we removed the label by filtering out all the points near the fitted plane. After the two-step filtering, a clean segmentation of the panicle point cloud can be retrieved (Figure 6c).
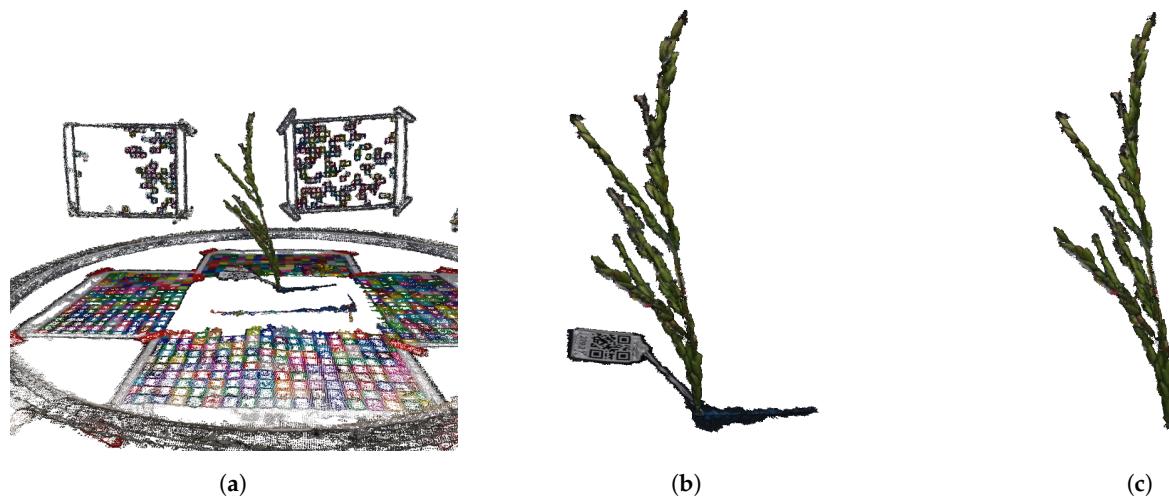

(**a**)   (**b**)   (**c**)

Figure 6. Point cloud post-process: (**a**) the de-noised point cloud generated by MVE; (**b**) the point cloud of the panicle with background removed; and (**c**) the final point cloud of the plant with labels removed.

3. Results

In this section, we conducted comparisons based on the results of our experiments, and the 3D models were illustrated as a mesh for better visualization. After the plants grew to a suitable height, we started to image them using our imaging systems periodically for the duration of the experiments. Then, these images were utilized to build 3D shapes using our reconstruction pipeline. With the 3D shapes, we can extract multiple phenotype

traits, such as leaf count, volume, and surface area [6,10,11]. In this work, the length of the panicles was utilized as the trait for results verification. The imaging process took up to two minutes to capture a set of images for one plant. Therefore, once optimized, the systems had the potential for high-throughput phenotyping.

3.1. Results Verification

To verify the results, we first performed a correlation analysis on the phenotype traits estimated based on the ground truth and the 3D reconstructed models. The ground truth was obtained by manually measuring the panicle lengths after harvest. The estimated lengths were obtained using the *Measuring Tool* application in MeshLab [31,32] to measure the lengths of our corresponding 3D reconstructions of panicles. The estimated lengths were then rescaled using the ring apparatus. Since we already obtained the physical size of the ring apparatus, the estimated lengths in the physical unit can be computed. The accuracy and the error were assessed using coefficient of determination (R^2) and mean absolute error (MAE), respectively.

As shown in Figure 7, 36 panicles samples were utilized for verification. The R^2 was 0.911, which indicated a high correlation between the model-derived length values and the manually measured values. The MAE was 1.05 cm, and it represented an error rate of 5.8% of the averaged panicle length given that the averaged panicle length was 18.15 cm in the experiment. The low MAE also implied a high accuracy of the estimated lengths and the high quality of the 3D models.
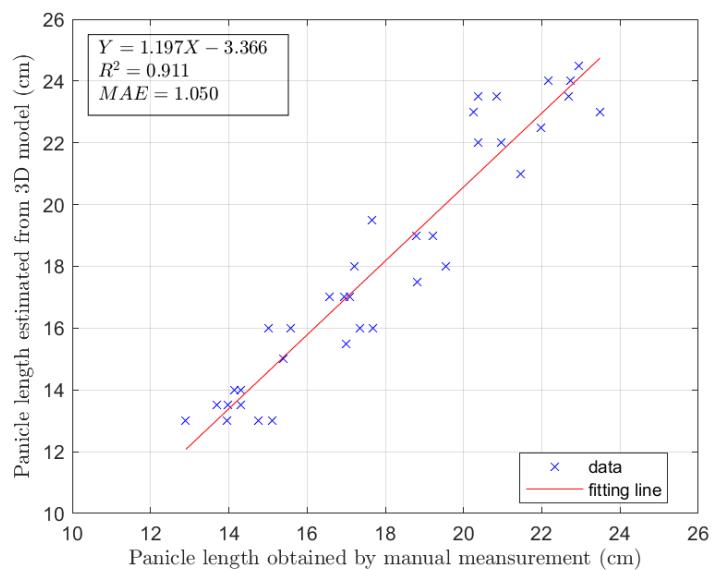

Figure 7. Panicle lengths obtained by estimation from 3D model vs. manual measurement.

3.2. Comparison of 3D Models with Various Number of Images and Cameras

The second experiment was conducted to evaluate the effect of the number of images and cameras on the reconstruction process. The computing platform we used was a computer with an Intel Core i7-8700 K CPU @3.70 GHz (Intel Co., Santa Clara, CA, USA) and 16 GB DDR4 random-access-memory.

As demonstrated in Table 1, Figures 8 and 9, we built a 3D model of maize as an example with several images and cameras (15, 20, 30, 60, or 120 images with one or two cameras). Inspired by the methods proposed by Lehtola et al. [33], we conducted a subjective assessment to evaluate the quality of the point clouds with metrics such as completeness and number of outliers. We found that at least 60 images were needed to build a good-quality 3D model, as shown in Figure 8a. As the number of images decreases, an obvious loss of quality of the 3D model can be observed. As illustrated in Figure 8b, there were holes on leaves in marked region 3, and part of the leaf was missing in marked

region 1 and 2 if only 30 images were used. This could possibly be due to a significant difference between sequential images when the total number of images was limited. A lack of matched image features between paired images can lead to an insufficient number of correspondences through all the images and, thus, an incomplete result. Moreover, as illustrated in the first three rows in Table 1, if the number of images was too low (lower than 30 in this case), MVE would fail to reconstruct 3D models since it cannot detect enough correspondences to generate 3D points. On the other hand, a higher number of images would not necessarily lead to a better result. Although the number of points in generated 3D models increased (shown in the last row in Table 1), the quality of the 3D model was not improved. For example, by comparing with the plants, we found fake branches in the reconstructed 3D model, as demonstrated in the marked region in Figure 8c. The reason for these fake branches was that the noise was erroneously considered as part of the stem. Additionally, when the number of images was too high, the computing time cost would increase dramatically, as shown in the last row in Table 1.

Table 1. Results with Various Numbers of Images and Cameras.

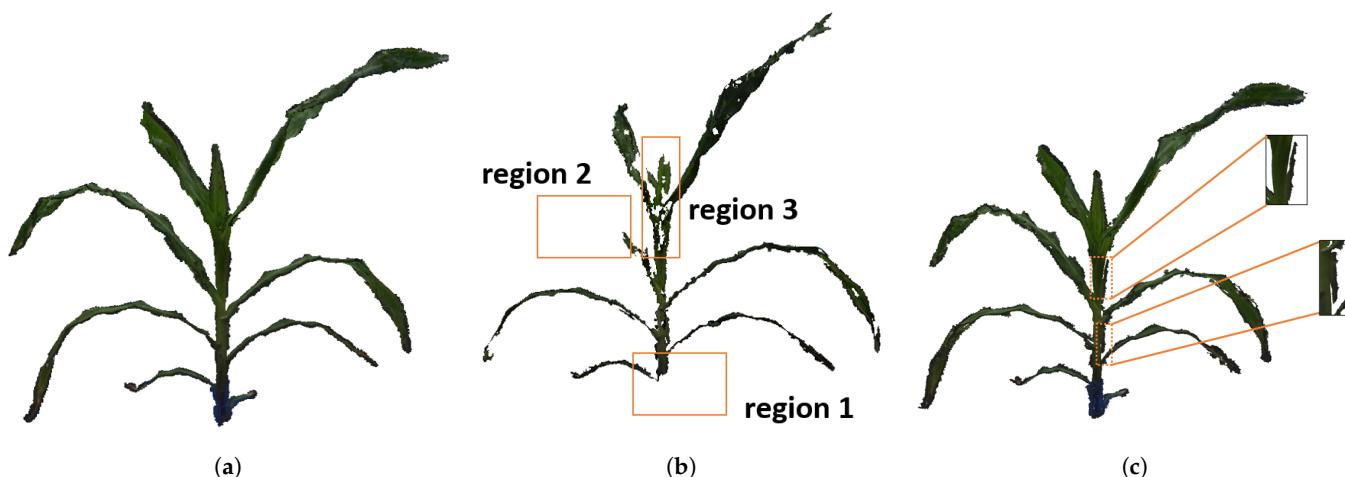| #Images (by Cam. 1)	#Images (by Cam. 2)	#Images in Total	Success in 3D Model Generation	#Point in Generated 3D Model	Time Cost (sec.)
15	0	15	No	N/A	N/A
10	10	20	No	N/A	N/A
15	15	30	No	N/A	N/A
30	0	30	Yes	45,881	391
30	30	60	Yes	98,101	1171
60	0	60	Yes	96,766	1300
60	60	120	Yes	129,657	2758

Figure 8. The generated models of maize with a various number of images: (**a**) a reconstructed model with 60 images (two cameras with 30 images per camera); (**b**) a reconstructed model with 30 images; and (**c**) a reconstructed model with 120 images.

By comparing the results with the same number of images, we also discovered that increasing the number of cameras would lead to the improvement of 3D models, although the number of points in generated 3D models and time cost made no differences as shown in the fifth and sixth rows in Table 1. Figure 9 shows a result comparison between the models using 60 images captured by two cameras (i.e., 30 images per camera) and one camera. It can be observed that the 3D model using two cameras (Figure 9a) had better quality than the one using one camera (Figure 9b), especially in the branches (the marked region in the figure). One possible reason could be that the cameras with various heights and view-angles reduced the occlusion and hence provide more correspondences and enhanced

3D reconstruction results. Therefore, there would be fewer missing points in the 3D models, especially in the region that can be easily occluded by the leaves or stem (such as branches).

Figure 9. The generated models of maize with a various number of cameras: (**a**) reconstructed model with 60 images taken by two cameras (i.e., 30 images per camera); and (**b**) reconstructed model with 60 images taken by one camera.

3.3. Evaluation of Color Checkerboards

One of the main differences between our imaging systems and existing ones was the usage of checkerboards. In this section, the importance of these checkerboards was evaluated with respect to image features and generated models.

3.3.1. Evaluation with Respect to Image Features

Image features were crucial for finding pixel pairs among images for 3D reconstruction. As a result, the quality of the reconstructed 3D shape would be greatly enhanced if the number of detected features was increased for each image. To evaluate the effect of color checkerboards with respect to image features, we captured panicle images from a similar view-angle with and without the checkerboards. As shown in Figure 10, the detected image features were visualized as red points (detected by SIFT) and green points (detected by SURF). By utilizing checkerboards in the imaging system, the number of detected image features increased from 2289 (1602 SIFT features and 687 SURF features) to 4141 (2690 SIFT features and 1451 SURF features). A higher number of image features facilitated us to generate more accurate parameters (e.g., camera parameters) in reconstruction.
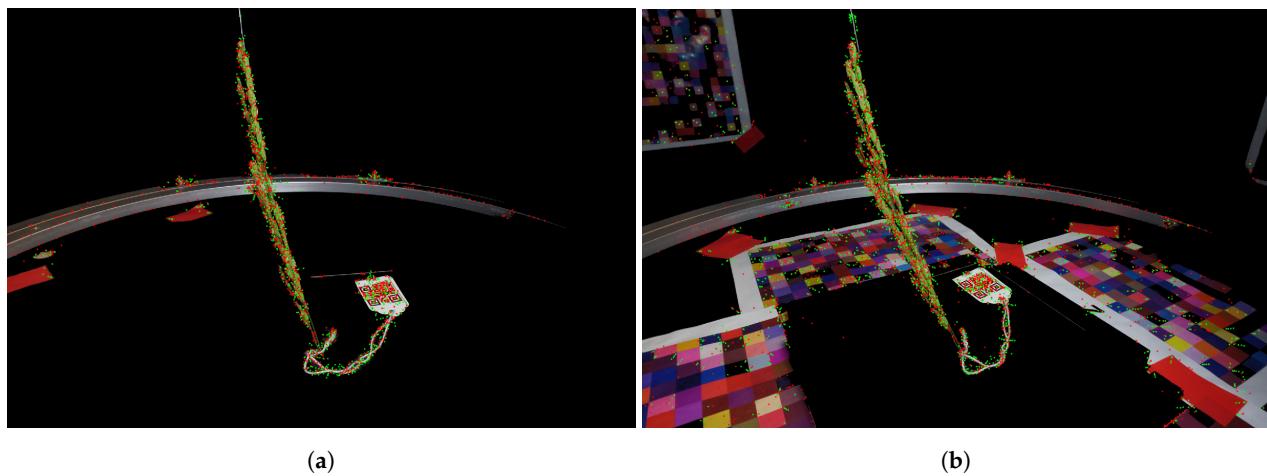

Figure 10. An example of detected imaging features: (**a**) In an image without checkerboards, 2289 features were found in total (1602 SIFT features and 687 SURF features). (**b**) In an image with checkerboards, 4141 features were found in total (2690 SIFT features and 1451 SURF features). Red points and green points indicate image features detected by SIFT and SURF, respectively.

3.3.2. Evaluation with Respect to Models

To further evaluate the effect of checkerboards, we also examined the models reconstructed by applying the same pipeline with and without checkerboards. In Figure 11, the first and second rows show the images and 3D shapes with and without checkerboards, respectively. By comparing the results generated using the same number of images, it was evident that the absence of checkerboards reduces the 3D reconstruction quality. Compared to the model generated from images with checkerboards (Figure 11b), the one without checkerboards included several missing parts in the marked regions (Figure 11d).
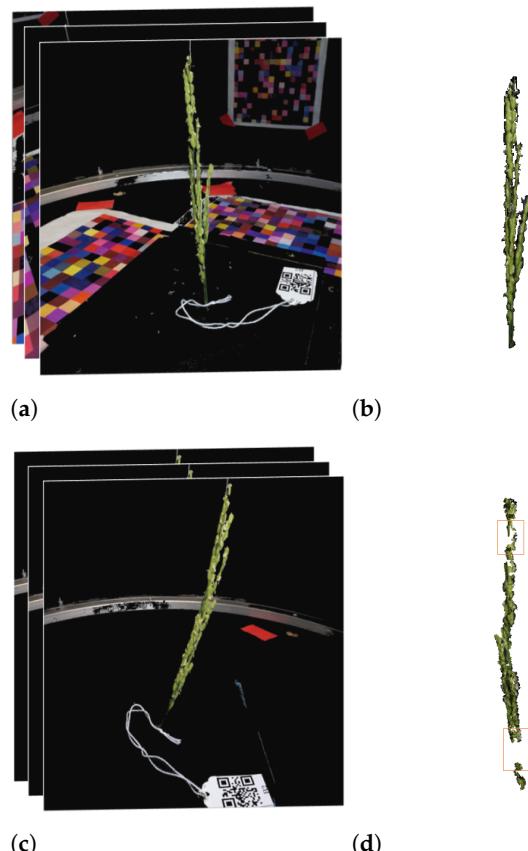

(**a**)   (**b**)

(**c**)   (**d**)

Figure 11. An example of input images and generated models: (**a**) input images with checkerboards; (**b**) generated models using 60 images from (**a**); (**c**) input images without checkerboards; and (**d**) generated models using 60 images from (**c**).

3.4. Evaluation of Stability

Another essential improvement of our systems was that we increased the system stability by rotating cameras rather than plants. In our systems, the positions of 3D points to be reconstructed were stationary, and the shutter speed was fast enough to eliminate possible camera instabilities incurred by the rotation of cameras. Therefore, the quality of the reconstructed models was improved. In this section, the importance of stability was evaluated by comparing the reconstructed models using our systems to the ones using the traditional turntable-based imaging system where plants were continuously rotated. The experiments were conducted on both maize and rice plants. For maize, both of the imaging systems were capable of generating models, as shown in Figure 12. However, the 3D shapes generated by the traditional turntable-based imaging system were consistent of lower quality for non-rigid plant parts in our experiments. Due to the motion of the plants, tissues such as leaves vibrated. The random vibration may lead to duplicated points in the final results. Figure 12b shows the model reconstructed using the traditional system, which included duplicated parts of leaves (e.g., the part in the marked region of Figure 12b). As

shown in Figure 12a, this issue was addressed in the result generated from our system because of the detaching of the plants and the ring apparatus.

For rice, we attempted to use the traditional system to reconstruct the whole plant and then retrieve the panicle segmentation for comparison. However, the models cannot be generated due to the complex plant architecture and higher vibrations of the long and flexible rice leaves and panicles when rotating the plant using the traditional turntable-based imaging system. In contrast, our system was able to generate high-quality 3D models for panicles, as shown in Figure 11b.
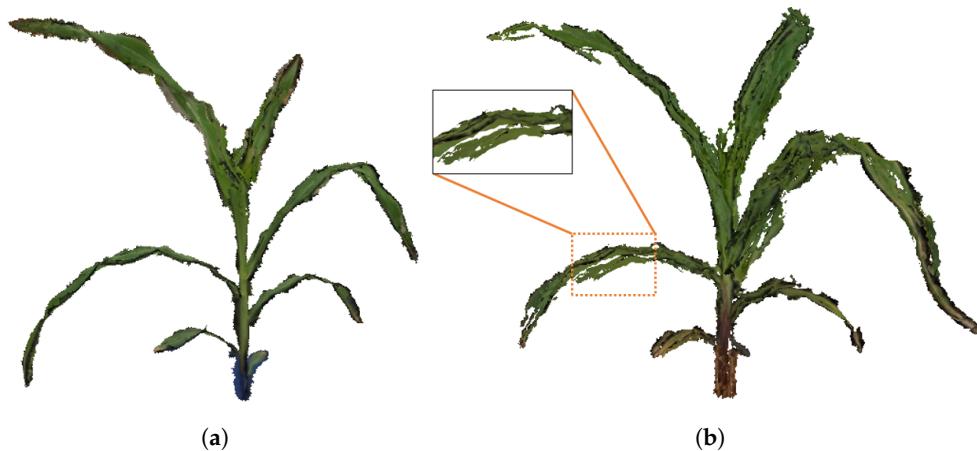

(a)   (b)

Figure 12. An example of models of maize from two imaging systems: (**a**) the model from our imaging system; and (**b**) the model from the traditional turntable-based imaging system.

4. Discussion

Though our imaging systems have demonstrated their potential to generate a more accurate point cloud than a traditional turntable-based imaging system, there is still space for improvement. First, our imaging systems are designed only for indoor experiments to avoid external noise from the environment. If performed in the field, the experiments will be affected by ambient conditions (e.g., wind) inevitably. As a result, the plants will probably vibrate in the process of imaging. Since stability is one of the essential requirements in our experiments, the movement caused by wind may lead to a noisy reconstructed point cloud or even a failure to generate a 3D model. Moreover, sunlight is also a problem in the field because the illumination on the object is constantly varying, caused by directional lighting and shading conditions. Therefore, it is not practical to set a constant threshold for filtering in the pipeline. The second area that could be enhanced is the reconstruction algorithm (MVE). In this work, there was no assumption made about the shape of the plant. In other words, the optimal species-specific priors were not developed. If this domain knowledge (e.g., leaf shapes and panicle structures) is utilized, the accuracy of 3D reconstruction can be enhanced. Third, as the number of images increases, the computation time complexity becomes a pipeline limitation for high-throughput applications. Since our current pipeline is CPU intensive, one of the possible solutions is to utilize GPU computing. Because some of the time-consuming steps, such as calculating coordinates of points in 3D space using matched image features, are parallelizable, it is possible to reduce the computation time cost significantly by running the pipeline on GPUs. Fourth, though we found that the reconstruction results were not sensitive to view-angle if two cameras were used, we did not thoroughly study the best view-angle selection of the cameras. We plan to conduct it with different plant geometrical structures in our future work.

5. Conclusions

In this work, we presented two imaging systems for plants of various sizes, as well as an end-to-end pipeline to reconstruct the 3D models. Our experimental set-up and

pipelines overcome several limitations of existing imaging systems and have the potentials for enhancing 3D high-throughput phenotyping. In both systems, plants remain still in the center, and the cameras rotate around the plants for stability. We also designed color checkerboards to provide additional image features that improve the accuracy of the reconstruction. In our experiments, we discussed how the number of images, number of cameras, and extra image features provided by checkerboards affect the generated 3D models. By comparison of the results from our systems and a traditional turntable imaging system, we illustrated the importance of plant stability. In summary, the proposed imaging systems can be directly used to reconstruct accurate 3D models of plants. For designers of new imaging systems, we provide our recommendations for various settings, such as checkerboards, plant stability, and multiple cameras, to improve accuracy of 3D reconstruction results. In the future, we plan to build a portable version of the imaging systems in a chamber to tackle wind and sunlight under outdoor conditions. We also would like to use species-specific priors to enhance the performance of the pipeline, and reduce the computation time complexity by developing a GPU-based pipeline.

References

1. Chew, Y.H.; Smith, R.W.; Jones, H.J.; Seaton, D.D.; Grima, R.; Halliday, K.J. Mathematical models light up plant signaling. *Plant Cell* **2014**, *26*, 5–20. [CrossRef] [PubMed]
2. Furbank, R.T.; Tester, M. Phenomics–technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* **2011**, *16*, 635–644. [CrossRef] [PubMed]
3. Zhou, Y.; Srinivasan, S.; Mirnezami, S.V.; Kusmec, A.; Fu, Q.; Attigala, L.; Fernandez, M.G.S.; Ganapathysubramanian, B.; Schnable, P.S. Semiautomated feature extraction from RGB images for sorghum panicle architecture GWAS. *Plant Physiol.* **2019**, *179*, 24–37. [CrossRef] [PubMed]
4. Gage, J.L.; Miller, N.D.; Spalding, E.P.; Kaeppler, S.M.; de Leon, N. TIPS: A system for automated image-based phenotyping of maize tassels. *Plant Methods* **2017**, *13*, 21. [CrossRef]
5. Klukas, C.; Chen, D.; Pape, J.M. Integrated analysis platform: An open-source information system for high-throughput plant phenotyping. *Plant Physiol.* **2014**, *165*, 506–518. [CrossRef]
6. Thapa, S.; Zhu, F.; Walia, H.; Yu, H.; Ge, Y. A novel LiDAR-based instrument for high-throughput, 3D measurement of morphological traits in maize and sorghum. *Sensors* **2018**, *18*, 1187. [CrossRef]
7. Zhu, F.; Thapa, S.; Gao, T.; Ge, Y.; Walia, H.; Yu, H. 3D Reconstruction of Plant Leaves for High-Throughput Phenotyping. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 4285–4293.
8. Gibbs, J.A.; Pound, M.; French, A.P.; Wells, D.M.; Murchie, E.; Pridmore, T. Approaches to three-dimensional reconstruction of plant shoot topology and geometry. *Funct. Plant Biol.* **2017**, *44*, 62–75. [CrossRef]
9. Fuhrmann, S.; Langguth, F.; Moehrle, N.; Waechter, M.; Goesele, M. MVE-An image-based reconstruction environment. *Comput. Graph.* **2015**, *53*, 44–53. [CrossRef]
10. Sandhu, J.; Zhu, F.; Paul, P.; Gao, T.; Dhatt, B.K.; Ge, Y.; Staswick, P.; Yu, H.; Walia, H. PI-Plat: A high-resolution image-based 3D reconstruction method to estimate growth dynamics of rice inflorescence traits. *Plant Methods* **2019**, *15*, 162. [CrossRef]

11. Gao, T.; Sun, J.; Zhu, F.; Doku, H.A.; Pan, Y.; Walia, H.; Yu, H. Plant Event Detection from Time-Varying Point Clouds. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3321–3329.

12. McCormick, R.F.; Truong, S.K.; Mullet, J.E. 3D sorghum reconstructions from depth images identify QTL regulating shoot architecture. *Plant Physiol.* **2016**, *172*, 823–834. [CrossRef]

13. Khoshelham, K.; Elberink, S.O. Accuracy and resolution of Kinect depth data for indoor mapping applications. *Sensors* **2012**, *12*, 1437–1454. [CrossRef]

14. Cao, W.; Zhou, J.; Yuan, Y.; Ye, H.; Nguyen, H.T.; Chen, J.; Zhou, J. Quantifying Variation in Soybean Due to Flood Using a Low-Cost 3D Imaging System. *Sensors* **2019**, *19*, 2682. [CrossRef]

15. He, J.Q.; Harrison, R.J.; Li, B. A novel 3D imaging system for strawberry phenotyping. *Plant Methods* **2017**, *13*, 93. [CrossRef]

16. Chaudhury, A.; Barron, J.L. Machine vision system for 3D plant phenotyping. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2018**, *16*, 2009–2022. [CrossRef]

17. Nguyen, T.; Slaughter, D.; Max, N.; Maloof, J.; Sinha, N. Structured light-based 3D reconstruction system for plants. *Sensors* **2015**, *15*, 18587–18612. [CrossRef]

18. Wu, S.; Wen, W.; Xiao, B.; Guo, X.; Du, J.; Wang, C.; Wang, Y. An accurate skeleton extraction approach from 3D point clouds of maize plants. *Front. Plant Sci.* **2019**, *10*. [CrossRef]

19. Zhang, Y.; Teng, P.; Shimizu, Y.; Hosoi, F.; Omasa, K. Estimating 3D leaf and stem shape of nursery paprika plants by a novel multi-camera photography system. *Sensors* **2016**, *16*, 874. [CrossRef]

20. Sony.com. Time-lapse (PlayMemories Camera App) | Sony USA. 2020. Available online: https://www.sony.com/electronics/support/articles/00021763 (accessed on 24 July 2012).

21. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]

22. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]

23. Fuhrmann, S.; Goesele, M. Floating scale surface reconstruction. *ACM Trans. Graph. (ToG)* **2014**, *33*, 1–11. [CrossRef]

24. Andrade, G.; Ramos, G.; Madeira, D.; Sachetto, R.; Ferreira, R.; Rocha, L. G-DBSCAN: A GPU accelerated algorithm for density-based clustering. *Procedia Comput. Sci.* **2013**, *18*, 369–378. [CrossRef]

25. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* **2002**, *80*, 76–87. [CrossRef]

26. Hunt, E.R., Jr.; Doraiswamy, P.C.; McMurtrey, J.E.; Daughtry, C.S.; Perry, E.M.; Akhmedov, B. A visible band index for remote sensing leaf chlorophyll content at the canopy scale. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 103–112. [CrossRef]

27. Wijayanto, A.W.; Triscowati, D.W.; Marsuhandi, A.H. Maize field area detection in East Java, Indonesia: An integrated multispectral remote sensing and machine learning approach. In Proceedings of the 2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, 6–8 October 2020; pp. 168–173.

28. Eng, L.S.; Ismail, R.; Hashim, W.; Baharum, A. The use of VARI, GLI, And VIgreen formulas in detecting vegetation in aerial images. *Int. J. Technol.* **2019**, *10*, 1385–1394. [CrossRef]

29. de Souza, E.G.; Scharf, P.C.; Sudduth, K.A. Sun position and cloud effects on reflectance and vegetation indices of corn. *Agron. J.* **2010**, *102*, 734–744. [CrossRef]

30. Andrade, R.G.; Hott, M.C.; Magalhães Junior, W.C.P.d.; D'Oliveira, P.S. Monitoring of Corn Growth Stages by UAV Platform Sensors. *Int. J. Adv. Eng. Res. Sci.* **2019**, *6*, 54–58. [CrossRef]

31. Cignoni, P.; Callieri, M.; Corsini, M.; Dellepiane, M.; Ganovelli, F.; Ranzuglia, G. Meshlab: An open-source mesh processing tool. In Proceedings of the Eurographics Italian Chapter Conference, Salerno, Italy, 2–4 July 2008; Volume 2008, pp. 129–136.

32. Callieri, M.; Ranzuglia, G.; Dellepiane, M.; Cignoni, P.; Scopigno, R. Meshlab as a complete open tool for the integration of photos and colour with high-resolution 3D geometry data. In Proceedings of the CAA2012 40th Conference in Computer Applications and Quantitative Methods in Archaeology, Southampton, UK, 26–30 March 2012; Volume II, pp. 406–416.

33. Lehtola, V.V.; Kaartinen, H.; Nüchter, A.; Kaijaluoto, R.; Kukko, A.; Litkey, P.; Honkavaara, E.; Rosnell, T.; Vaaja, M.T.; Virtanen, J.P.; et al. Comparison of the selected state-of-the-art 3D indoor scanning and point cloud generation methods. *Remote Sens.* **2017**, *9*, 796. [CrossRef]