Field-based Robotic Phenotyping of Sorghum Plant Architecture using Stereo Vision

Yin Bao[1], Lie Tang[1], Matthew W. Breitzman[2], Maria G. Salas Fernandez[2], Patrick S. Schnable[2]

[1]Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, Iowa, 50011

[2]Department of Agronomy, Iowa State University, Ames, Iowa, 50011

Abstract

Sorghum (*Sorghum bicolor*) is known as a major feedstock for biofuel production. To improve its biomass yield through genetic research, manually measuring yield component traits (e.g. plant height, stem diameter, leaf angle, leaf area, leaf number, and panicle size) in the field is the current best practice. However, such laborious and time-consuming tasks have become a bottleneck limiting experiment scale and data acquisition frequency. This paper presents a high-throughput field-based robotic phenotyping system which performed side-view stereo imaging for dense sorghum plants with a wide range of plant heights throughout the growing season. Our study demonstrated the suitability of stereo vision for field-based 3D plant phenotyping when recent advances in stereo matching algorithms were incorporated. A robust data processing pipeline was developed to quantify the variations or morphological traits in plant architecture, which included plot-based plant height, plot-based plant width, convex hull volume, plant surface area, and stem diameter (semi-automated). These image-derived measurements were highly repeatable and showed high correlations with the in-field manual measurements. Meanwhile, manually collecting the same traits required a large amount of manpower and time

compared to the robotic system. The results demonstrated that the proposed system could be a promising tool for large-scale field-based high-throughput plant phenotyping of bioenergy crops.

KEYWORDS

high-throughput phenotyping, stereo vision, point cloud processing, bioenergy sorghum, yield component traits

1. INTRODUCTION

Sorghum (*Sorghum bicolor*) is regarded as one of the most productive bioenergy crops in the US because of its drought tolerance and high biomass yield potential (Rooney, Blumenthal, Bean, & Mullet, 2007). Increasing biomass yield is a major objective for biomass sorghum breeding programs. However, breeding high-yield crop varieties is a difficult task because biomass yield is a complex trait that depends on many different yield component traits. It has been shown that plant architectural traits such as plant height (Salas Fernandez, Becraft, Yin, & Lübberstedt, 2009), leaf area index (Olson et al., 2012), leaf angle (Truong, McCormick, Rooney, & Mullet, 2015) are highly correlated with sorghum biomass yield. Unraveling the genetic control of such plant architectural traits at different growth stages can potentially accelerate plant breeding programs. The advances in high-throughput genotyping technologies have allowed for easy access to the massive plant genomic data. However, linking phenotypic traits to its underlying genes often requires accurately measuring the traits for many genotypes and replications. For instance, genome-wide association studies (GWAS) are routinely used nowadays to efficiently identify the genetic regions controlling multiple traits at the same time from a diverse population (Zhao, Mantilla Perez, Hu, & Salas Fernandez, 2016). However, the most commonly used practice to obtain the desired phenotypic data is in-field manual measurement. Such tasks are

extremely laborious and time-consuming, considerably limiting experiment scale and data acquisition frequency. High-throughput field-based plant phenotyping aims at resolving this bottleneck.

In the last decade, many automated high-throughput field-based phenotyping platforms have emerged. Several ground-based systems were designed to travel over short row crops and carry sensors such as RGB, depth or hyperspectral cameras, light curtains, and infrared radiometers (Andrade-Sanchez et al., 2014; Jiang et al., 2018; Ruckelshausen et al., 2009; Shafiekhani, Kadam, Fritschi, & DeSouza, 2017; Underwood, Wendel, Schofield, McMurray, & Kimber, 2017). Plant population, canopy density, inter-row spacing, plant height, leaf area, light exposure, stem thickness, canopy temperature, and canopy reflectance were measured. Designing a high-clearance vehicle to characterize sorghum plant architecture is a challenging task because some varieties can grow up to 4 m. The Field Scanalyzer (LemnaTec, Aachen, Germany) is a stationary gantry system which can move a sensor platform from 3 m to 6 m above the ground (Virlet, Sabermanesh, Sadeghi-Tehran, & Hawkesford, 2017). The major limitations of this approach are the reduced field size, fixed location, and high cost. Unmanned aerial imaging systems have been developed for high-throughput field phenotyping (Chapman et al., 2014; Shi et al., 2016; Watanabe et al., 2017). Regarding plant architectural traits, plant height and leaf area index were extracted so far because top-view aerial imaging could only capture top-level visible canopy. Recently, two ground-based robotic systems were developed specifically for phenotyping bioenergy sorghum. One, named Robotanist, autonomously navigated in the 0.61 m row space between plots (Mueller-Sim, Jenkins, Abel, & Kantor, 2017). It had a 1.83 m vertical linear track to move a robotic manipulator for stem strength measurement, which was guided by a side-viewing stereo camera mounted next to the robotic gripper. The other system traveled

inside the 1.8 m alley way between sub plots (Vijayarangan et al., 2018). Two foldable sensor booms could move vertically on a mast higher than 4 m while four sensorpods on the booms continuously captured the side-view images of plants with multiple cameras of different resolutions and wavelengths.

Digital plant architecture phenotyping relies on the three-dimensional (3D) reconstruction of plants. Although different types of 3D sensing technologies have been investigated for plant phenotyping, very limited 3D sensors are suitable for in-field sorghum phenotyping. Time-of-Flight cameras often do not perform well under direct sunlight, and they were mostly used for plant phenotyping in controlled environments (Chaivivatrakul, Tang, Dailey, & Nakarmi, 2014; Jiang, Li, & Paterson, 2016; J. Li & Tang, 2017). Line-scan LiDAR sensors can work outdoor, and were used for leaf area density estimation (Sanz, Rosell, Llorens, Gil, & Planas, 2013), individual tree identification (Underwood, Jagbrant, Nieto, & Sukkarieh, 2015), and tree trunk detection (Bargoti, Underwood, Nieto, & Sukkarieh, 2015). However, due to the low spatial resolutions, thin structures such as stems are difficult to reconstruct accurately with LiDAR sensors. Stereo vision has shown promising potentials in plant phenotyping applications. Active stereo vision with flash illumination was used to detect round fruits of textureless surfaces such as apple and grape under field conditions (Pothen & Nuske, 2016). The high-powered flash illumination combined with a fast shutter speed produced good image quality under sunlight. The same imaging system was adopted on the Robotanist (Mueller-Sim et al., 2017) for sorghum stem count and stem width estimation (Baweja, Parhar, Mirbod, & Nuske, 2018). In controlled environments, multi-view stereo can provide a detailed 3D reconstruction at a low cost (Lou, Liu, Han, & Doonan, 2014; Nguyen, Slaughter, Maloof, & Sinha, 2016). Structured-light pattern illumination was used to add texture on the textureless plant surfaces to facilitate stereo matching

(Nguyen et al., 2016). The main limitation of multi-view stereo reconstruction is the long processing time. Recently, multi-view stereo with active illumination was also applied to reconstruct a side-view 3D point cloud of fully grown sorghum plants in the field (Sodhi, Vijayarangan, & Wettergreen, 2017). However, since the cameras were moved to scan the tall plants, wind could be a potential problem. Wind can cause large movements of plant canopies, especially for biomass sorghum with long and broad leaves. If the plants were moving during the scanning process, the 3D reconstruction integrity would be corrupted.

This paper presents a high-throughput field-based phenotyping system for the characterization of plant architecture of dense canopy sorghum solely relying on stereo vision. Instead of multi-view stereo reconstruction using moving cameras, our solution was to image the potentially fast moving plants with multiple stereo cameras simultaneously and instantaneously (using a fast shutter speed). We demonstrate that the depth sensing capability of conventional passive stereo vision can be quite robust against field lighting conditions with the state-of-the-art stereo matching techniques. Furthermore, several valuable features were extracted from stereo 3D reconstruction to quantify some important plant architectural variations including plot-based plant height, plot-based plant width, convex hull volume, plant surface area, and stem diameter. These features demonstrate the advantages of the side-view imaging configuration for exposing plant architecture. The system was successfully used to collect and process image data for large-scale field experiments multiple times throughout the growing season. Discussion is provided about the results and limitations regarding the image data acquisition system, the stereo reconstruction algorithms, and the extracted features, followed by potential solutions to each of the three aspects for future improvements.

The objectives of this research project were to: (1) develop a field-based robotic ground vehicle to perform stereo imaging of side-view profile for biomass sorghum of various heights, (2) develop a robust 3D feature extraction pipeline to quantify variations in plant architecture, (3) analyze the repeatability of the extracted features and their correlations with the manual measurements, and (4) analyze the effects of different stereo matching algorithms on feature extraction results.

The paper is organized as follows: The field-based robotic ground vehicle is described in Section 2. Section 3 presents the data processing pipeline, including stereo reconstruction and feature extraction. In Section 4, the extracted features are analyzed with respect to the manually measured phenotypic traits, followed by a detailed time analysis. Section 5 provides discussions about the performance, the limitations, and the potential improvements for the data acquisition system, the stereo reconstruction algorithm, and the traits extraction, concluding in Section 6.

2. THE AUTONOMOUS DATA ACQUISITION SYSTEM (PHENOBOT 1.0)

A stereo image data acquisition system was designed for tall biomass crops, high-throughput data collection, and high-endurance operation. It was called Phenobot 1.0. This section starts with the field design and imaging scheme, followed by a detailed description of the components and capabilities of Phenobot 1.0. The data acquisition workflow and performance is presented at the end.

2.1. Field Design and Imaging Scheme

Field experiments were conducted during the summers of 2014 and 2016 at the Agricultural Engineering and Agronomy Research Farms of Iowa State University. In 2014, 589 diverse sorghum genetic lines were planted at two locations with two replications per location. The field

design and imaging scheme are illustrated in Figure 1. Each genotype was planted in a three-meter-long two-row plot with an average interplant spacing of 0.075 m. The row spacing within a plot was 1.5 m, while the row spacing between plots, where the robot navigated, was 2.2 m. Each side of a plot was imaged once per pass, with an imaged plot length of 1.5 m at the center of the plot. The wide row spacing was designed to prevent long leaves blocking the camera view. The effect of such wide row spacing is discussed in Section 5. In 2016, a subset of six sorghum lines representing the genetic diversity of the entire set were planted at one location with three repetitions for the ground-truth validation. The field design was the same as the one used for the entire set of diverse lines in 2014.
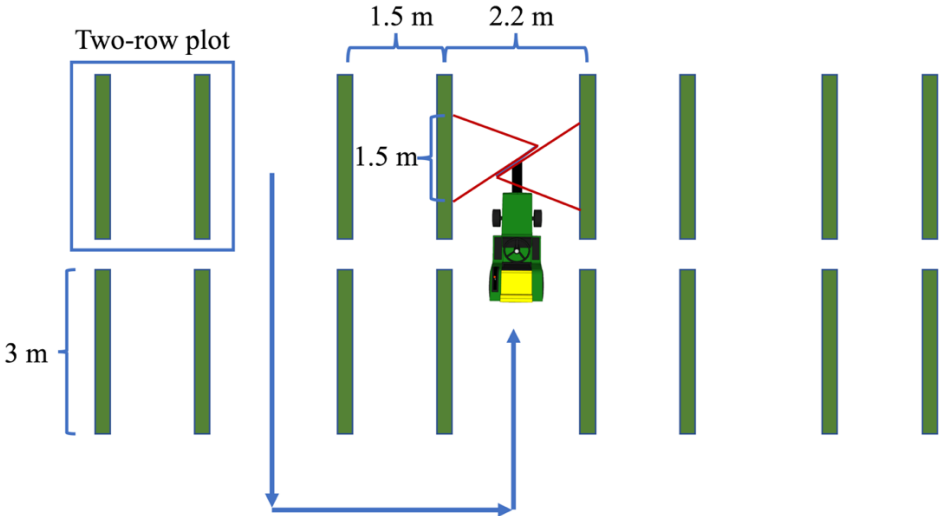

Figure 1: Field design and imaging scheme in a bird's-eye view. Each genotype is planted in a three-meter-long and two-row plot. The row spacing within a plot is 1.5 m, while the row spacing between plots is 2.2 m. The robot navigates between plots, and the imaging system has a horizontal field of view of 1.5 m. Two rows are imaged per pass.

2.2. Phenobot Development and Construction

The field-based stereo imaging robot consisted of four major components: a mobile platform, a navigation system, a stereo imaging system, and a central computer for data acquisition (Fig. 2). The mobile platform was a John Deere 1026R sub-compact utility tractor (John Deere, Moline, IL), which offered hours of continuous operation in the field. To realize autonomous navigation, the tractor was retrofitted with a Topcon 350 auto-guidance system (Topcon, Tokyo, Japan). Its AGI-4 receiver/steering controller (Topcon, Tokyo, Japan) integrated both a real-time kinematic global positioning system (RTK-GPS) and a steering controller. The tractor steering wheel was replaced by an AES-25 electric steering wheel (Topcon). A user can record a path during manual driving via the X30 console (Topcon). Afterwards, the auto-guidance system performed autonomous steering to follow the path. This integration resulted in an approximately 2-cm lateral tracking error at speeds below 1 m/s when following straight paths in the field. The vehicle travel speed was controlled by a linear actuator which was connected to the chassis and the forward speed pedal with two rotational joints. The cylinder position was set via RS-232 serial communication. After plant emergence, the robot was manually driven through the crop rows while the paths were recorded for subsequent navigation during the remaining growing season. Since this robot was designed to image the center of each three-meter plot, the Universal Transverse Mercator (UTM) coordinates of the robot were recorded at every imaging location during the initial path recording. These UTM coordinates were converted from the GPS coordinates output by the auto-steering controller at 10 Hz, and used as a map for season-long use. Note that the mapping process was only done once per field.

The stereo imaging system consisted of 12 Point Grey GRAS-20S4C-C 2-Megapixel color cameras daisy-chained on an IEEE-1394b bus. These cameras were mounted on a rotatable rig in

front of the vehicle, and arranged in six side-viewing binocular stereo camera heads (Fig. 2 right). Three of the stereo camera heads aimed to image the left-side crop row, and the other three, the right-side, allowing for imaging two rows in one pass. The set of stereo camera heads on the same side of the rotatable rig was placed at three height levels, to reduce occlusion caused by the long leaves growing towards the row space. All cameras imaging the same side were synchronously triggered using its FlyCapture software development kit. The 32-MB onboard image buffer provided temporary storage in case that the images were queued before they were saved in the central computer. An Edmund Optics 6-mm f/1.4 lens with a 62.1° angular field of view (FOV) was used along with a linear polarizing filter to suppress glare from the leaf and stem surfaces. Given the field design, the three-level stereo imaging configuration achieved a horizontal FOV of approximately 1.5 m (Fig. 1) and a vertical FOV of 2.7 m.

Figure 2: Phenobot without a camera extension rig for early stage imaging (left). Phenobot with a camera extension rig for imaging tall plants (right).

2.3. Data Acquisition

The data acquisition software ran on a Getac B300 rugged laptop with an Intel Core i7-3520M processor clocked at 2.9 GHz. The program workflow is illustrated in Figure 3. During data acquisition, the main thread of the program monitored the distance between the robot and the next imaging location along the crop row direction. Once the robot reached the location, the cameras on the corresponding side were triggered, and proper file names with embedded crop row IDs and plot IDs were constructed and saved in a list. Two other worker threads kept polling the camera image buffers on the left side and the right side, respectively, fetched images and saved them with proper file names in two separate solid state drives (SSD). Phenobot 1.0 achieved a maximum data collection speed of 0.67 m/s without camera image buffer overflow. Given the maximum travel speed and plot length, the average data throughput was 5 MB/s.
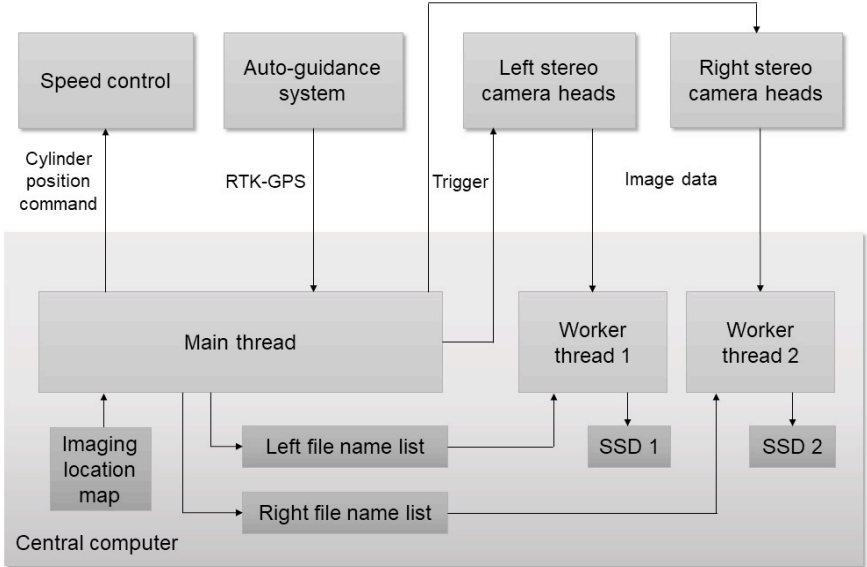

Figure 3: The multi-threaded data acquisition workflow. The main thread monitors robot location relative to the next image location in the map, triggers the left or right side stereo cameras, and inserts a new file name in the file name list. The left-side images are retrieved and saved by the

worker thread 1, while the right-side images are managed by the worker thread 2. The robot

speed is set by a user and handled by the main thread.

3. DATA PROCESSING

3.1. Dense Two-View Stereo 3D Reconstruction for Field Crops

Two-view stereo reconstruction is the process of transforming a stereo image pair into a 3D point

cloud. It consists of three steps: (1) stereo camera calibration and rectification, (2) stereo

matching, and (3) reprojection of the disparity map to a 3D point cloud. Step 1 and Step 3 are

well-established procedures nowadays, but the second step has the biggest impact on 3D

reconstruction quality, and remains an active research area in computer vision.

Our camera calibration process is described as follows. Each camera was first calibrated to

obtain its intrinsic parameters and distortion coefficients. Then, each stereo camera head was

calibrated and rectified such that the two image planes were the same and all epipolar lines were

parallel. Lastly, the relative pose (translation and rotation) between every two adjacent stereo

camera heads on the same side was calibrated. The wide-angle lens ensured sufficient overlaps

between the FOVs of the adjacent stereo camera heads at a close distance for the relative pose

calibration. The goal was to use the stereo camera pose at the bottom level as the world

coordinate system into which the reconstructed point clouds from the middle and top stereo

cameras were transformed. The stereo camera calibration and rectification were performed with

a checkerboard pattern using the OpenCV library (Bradski & Kaehler, 2000).

Dense two-view stereo matching is the process of correspondence search for every pixel between

the two stereo images, and relies on image texture. Field crops present several challenges for

dense two-view stereo correspondence matching. First, field lighting conditions can change

dramatically, which results in overexposed or underexposed images, diminishing image texture. Second, the nearly homogeneous color of leaf and stem surfaces can create large untextured regions in the images, adding more ambiguity to the stereo matching problem. Third, the dense canopy often causes heavy occlusion. The occluded pixels in the stereo images do not have correspondences. Fourth, the leaf and stem surfaces become non-Lambertian when the surface normal direction is close to being perpendicular to the image plane under strong sunlight, resulting in specular reflection and loss of image texture. Lastly, thin structures, such as stems, are often difficult to preserve in the stereo matching process.

Here we briefly review how different aspects of stereo matching algorithms were improved to solve these challenges, and we provide the rationale of the stereo matching algorithm chosen for this application. Most stereo matching algorithms consist of the following steps: matching cost computation, cost aggregation, disparity computation (and optimization), and disparity refinement (Scharstein & Szeliski, 2002). In matching cost computation, a 3D cost volume is built where each layer has the same dimensions as the input image and stores the matching costs for every pixel at some discretized disparity level. The performances of various conventional matching costs were evaluated on images with radiometric differences (exposure, vignetting, varying lighting, and noise) (Hirschmüller & Scharstein, 2009). Census was found to give the best overall performance (Zabih & Woodfill, 1994). With the recent breakthrough in artificial intelligence, a convolutional neural network (CNN) known as MC-CNN was trained to compute the similarity between two small image patches (Zbontar & LeCun, 2016). The architecture was a siamese network, which started with two convolutional feature extractors for the stereo image patches, and then concatenated the feature vectors and compared them with several fully connected layers. Surprisingly, MC-CNN learned to compare features invariant to exposure and

illumination differences. It was also insensitive to small stereo rectification errors, which are more present in high-resolution stereo images. Moreover, MC-CNN outperformed Census, especially on large untextured regions.

The matching cost of a single pixel can be ambiguous. Matching costs are often aggregated over a support window to improve matching accuracy. However, the side effect is blurred depth discontinuity. To preserve edges, a bilateral filter was used as the weighting function within the support window (Yoon & Kweon, 2006). Later, the geodesic distance support weight was adopted to further improve the edge preserving performance (Hosni, Bleyer, Gelautz, & Rhemann, 2009). In case of overlapping leaves, the weights drop sharply once beyond the contour of the center leaf due to the image edge and geodesic connectivity constraint. The above support windows all assume a constant disparity value for the pixels inside them, which is only true for planes parallel to the image plane. For slanted planes, they lead to stair-casing artifact. PatchMatch Stereo (PMS) solved this problem with slanted support windows in which disparities at different positions were parameterized by three plane coefficients (Bleyer, Rhemann, & Rother, 2011). One limitation of the window-based methods is that the fixed window size does not handle large untextured regions well. Efficient non-local cost aggregation was achieved by traversing the minimum spanning tree (MST) built on the stereo image (Yang, 2012). The support weights around a pixel are approximations of the geodesic support weights. Because the support region becomes the entire image, this method greatly improves matching robustness over large untextured regions. Cost aggregation can be viewed as filtering of the 3D cost volume (Hosni, Rhemann, Bleyer, Rother, & Gelautz, 2013).

To compute a disparity map, "winner-take-all" selects the disparity with the minimum matching cost for each pixel. However, advanced optimizations on cost volume have been studied to

enforce spatial smoothness constraint. The state-of-the-art smoothness priors regularize piece-wise affine planar surfaces (Olsson, Ulén, & Boykov, 2013; Ranftl, Bredies, & Pock, 2014). The optimization problem can be solved by graph cuts and variational methods at the expense of a high computational cost. Semi-Global Matching performs 1D scanline optimization in 16 directions of the 2D image space and sum up all the costs to enforce the smoothness constraint (Hirschmüller, 2008). Due to its efficiency and robustness, it has gained popularity in real-world applications.

For 3D reconstruction of in-field crops, the optimal stereo matching algorithm should be the one that can handle radiometric differences caused by field lighting conditions, reconstruct large untextured free-form surfaces, and preserve accurate depth discontinuity and thin structures. One of the top performing algorithms on Middlebury Stereo Evaluation 3.0 is 3DMST (Lincheng Li, Yu, Zhang, Zhao, & Zhang, 2017). This algorithm meets the above requirements by efficiently combining MC-CNN matching cost, PMS, and MST-based non-local edge-preserving support region filtering. Therefore, 3DMST was used as the main stereo matching algorithm in our study. Using our in-field sorghum dataset, a qualitative comparison between 3DMST and the OpenCV's Semi-Global Block Matching (SGBM) is illustrated in Figure 4. SGBM is a more efficient variant of the original SGM, and serves as a good baseline for the evaluation of 3DMST. The disparity maps and color-coded surface normal maps demonstrate 3DMST's superior edge-preserving piece-wise smoothness property for densely populated sorghum plants.
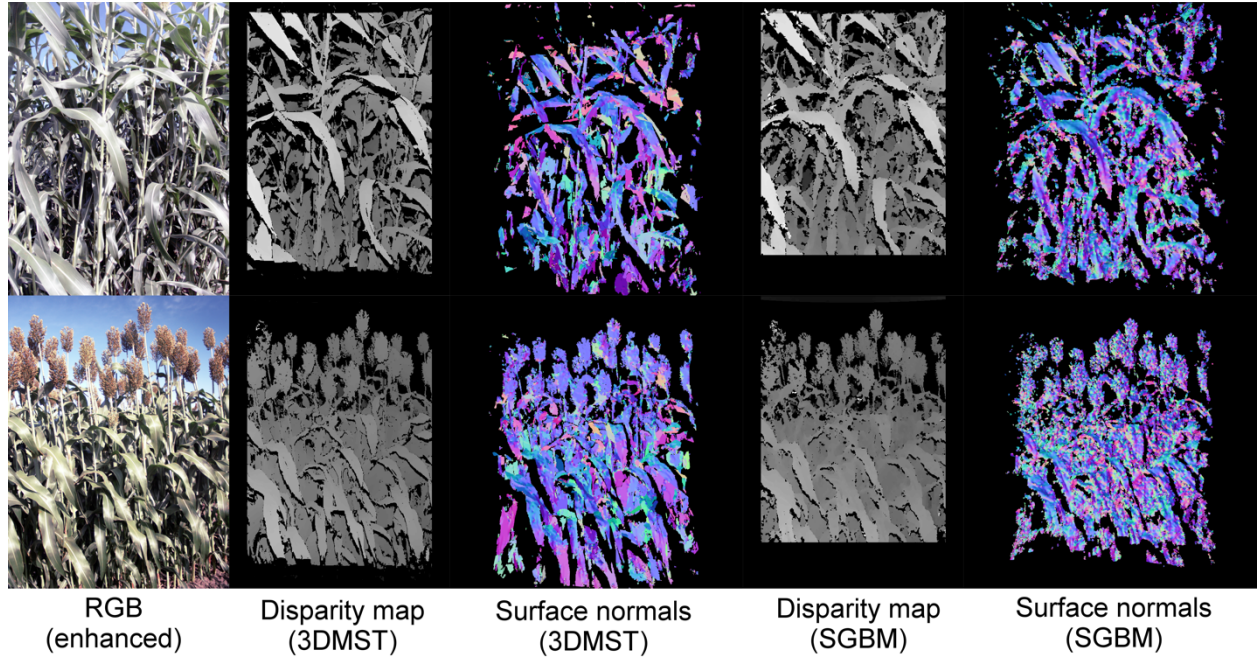
Figure 4: Qualitative comparison of 3D reconstruction quality between 3DMST and SGBM for in-field sorghum. The five images in each row correspond to an enhanced RGB image after histogram equalization, the disparity map of 3DMST, the colored-coded surface normal map of 3DMST, the disparity map of SGBM, and the colored-coded surface normal map of SGBM. 3DMST generates a more edge-preserving and piece-wise smooth surface reconstruction.

3.2. Point Cloud Preprocessing

The input 3D point cloud for feature extraction was the fusion of three point clouds from the three stereo camera heads on the same side of the rotatable rig. The processing pipeline was developed by utilizing the Point Cloud Library (PCL) (Rusu & Cousins, 2011). First, we defined our world coordinate system as follows: the X-axis was perpendicular to the ground plane pointing upwards, the Y-axis was parallel to the crop row, and the Z-axis was perpendicularly pointing towards the crop row. Due to the side-viewing angle, the crop rows beyond the foreground row were partially captured as well, and thus, background rows were removed by

filtering the *Z* coordinates. The threshold was set to 2.0 m, which equaled the distance between the bottom stereo camera and the midpoint between the foreground row and the first background row along the *Z*-axis. The resultant 3D point cloud was downsampled with the VoxelGrid (Rusu, 2010) filter of a voxel size of 1 cm, which served as an effective way to regularize point cloud density and speed up subsequent processing. A voxel can be viewed as a small cube. The point cloud was partitioned by a 3D voxel grid and all points in each voxel were replaced by their centroid. After this process, the point cloud only contained the foreground crop row and ground surface which was robustly identified by plane fitting with random sample consensus (RANSAC) (Fischler & Bolles, 1987). Because the ground plane was approximately perpendicular to the *X*-axis, the maximum deviation between the randomly sampled plane normal direction and the *X*-axis was limited to 20°. Any points within 3 cm away from the plane were considered as inliers. After the RANSAC plane fitting, the inliers of the ground surface were removed from the point cloud. X_{ground} was computed as the average of the *X* coordinates of the ground inliers to serve as the reference for measuring plant height. Next, the StatisticalOutlierRemoval (Rusu, Marton, Blodow, Dolha, & Beetz, 2008) filter was applied to further remove sparse outliers. This filter computes the distance from each point to all its neighbors and removes points whose mean distance to its neighbors are outside the interval defined by the global mean distance and standard deviation. The number of neighbors to analyze for each point was set to 50, and the standard deviation multiplier to 1. Last, the Euclidean clusters were extracted using a tolerance of 5 cm, and the small clusters of less than 400 points were removed to correct for the presence of weeds and soil clumps that failed the ground inlier test. The outcomes of the preprocessing stage are X_{ground} and the point cloud of the foreground crop row as illustrated in Figure 5.

Input: the point cloud of
stereo reconstruction

Background removal

VoxelGrid downsampling

RANSAC ground plane fitting
and ground removal

StatisticalOutlierRemoval filtering

EuclideanCluster filtering

Output: the point cloud
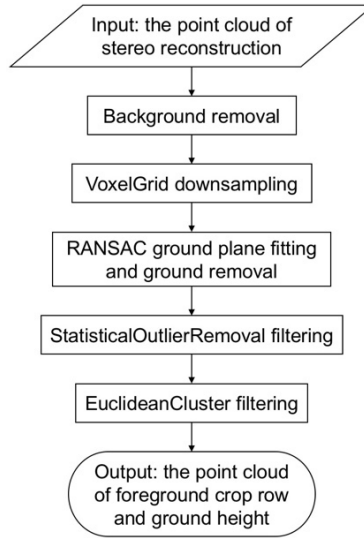of foreground crop row
and ground height

Figure 5: Workflow of the 3D point cloud preprocessing. The foreground crop row and ground surface were first extracted from the input point cloud. After the VoxelGrid downsampling, the ground plane was identified by random sample consensus (RANSAC). Sparse outliers and small clusters were then removed by the StatisticalOutlierRemoval filter and the EuclideanCluster filter, respectively. The preprocessing output is the point cloud of the foreground crop row and the ground height.

3.3. Plot-based Plant Height and Plot-based Plant Width

An axis-aligned bounding box (AABB) was extracted from the preprocessed point cloud. Each edge of an AABB is parallel to one of the axes of the world coordinate system. Therefore, an AABB was defined by two of its vertices, $P_{min}(X_{min}, Y_{min}, Z_{min})$ and $P_{max}(X_{max}, Y_{max}, Z_{max})$. P_{min} was obtained by finding the minimum coordinate in each dimension of a 3D point cloud. Similarly, P_{max} was constructed by the maximum coordinates. We defined plot-based plant height PPH $= X_{max} - X_{min}$ and plot-based plant width PPW $= Z_{max} - Z_{min}$ based on the world coordinate system described in previous sections.

However, to obtain more accurate and robust measurements, it is necessary to refine (X_{max}, X_{min}, Z_{max}, Z_{min}). Firstly, X_{min} is assigned to the value of X_{ground} considering that plant height should be measured from ground plane. Second, X_{max} is not robust against abnormal individuals which are much taller than average plants of a plot. Such phenomenon could be caused by the instability of a gene known to control sorghum plant height (Multani et al., 2003), by environmental variation affecting the germination rate within a plot, or by management practices such as unequal planting depth within a plot. Our solution was to partition the point cloud along crop row direction (Y-axis) into N_{slice} slices. In Slice i, let X_{max}^i denote the maximum X value of all points and w^i the associated weight equal to the ratio of number of points in Slice i to the total number of points. Weight w^i effectively reduces the contribution of slices that contain empty space. The weighted median of X_{max}^i was used as a robust reference of the top of a plot in case of abnormal plants. The weighted median was obtained by first sorting X_{max}^i and then finding the first X_{max}^k satisfying $\sum_{i=1}^{k} w^i \geq 0.5$. Then X_{max} of the AABB was refined by the weighted median value. The same weighted median approach was used to refine Z_{max} and Z_{min}. Note that the stereo cameras only captured one side of a crop row and thus, the extracted plant width represents half of the canopy depth along the Z-axis. Lastly, any point outside the refined AABB was removed.

The only parameter used to estimate PPH and PPW is the number of volume slices, N_{slice}. If N_{slice} equals one, the resultant PPH measures the maximum plant height, which may not be representative. The linear increase of N_{slice} reduces the thickness of each volume slice. In general, as the volume slice becomes thinner, the number of volume slices containing gaps between top canopies increases, resulting in a decrease in the estimated PPH (Fig. 6).
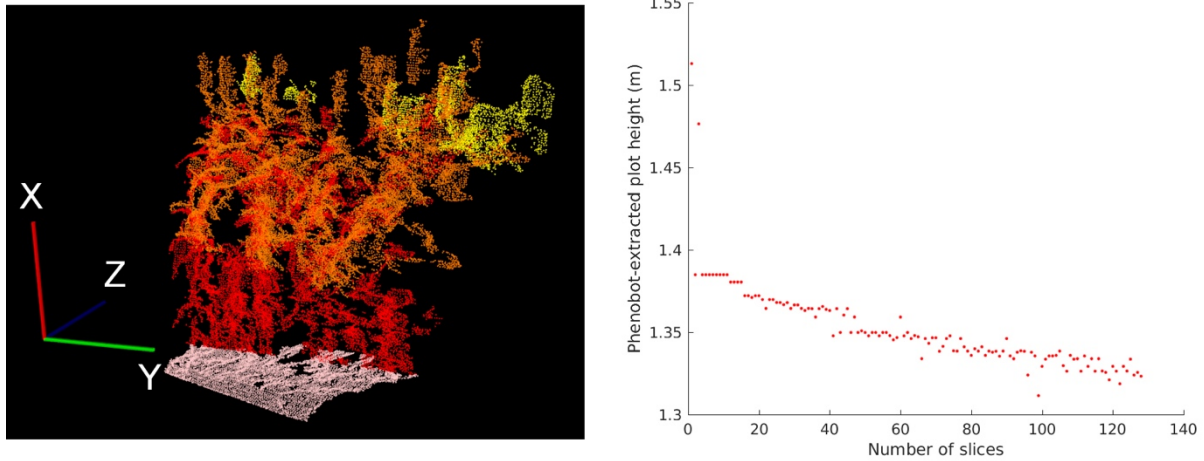
Figure 6: Sample 3D point cloud (left) and plot-based plant height vs number of volume slices (right). Left: different colors represent different point clouds of the ground surface and the three sets of stereo cameras.

The choice of a reasonable N_{slice} depends on plant growth stage and intra-row spacing. Before flowering, PPH is normally measured from ground to plant whorl. The top three or four leaves are frequently above the whorl. A large N_{slice} would help reduce the difference between actual plant height and our estimated PPH. After flowering, this potential inaccuracy does not need to be overcome since the exserted panicle appears above the leaves and its clearly exposed tip is used to estimate plant height. If N_{slice} is chosen so that only one plant is in each volume slice, our method would measure the exact PPH of each plant. However, choosing different N_{slice} values during the growing season is difficult because sorghum genotypes vary in flowering time. Therefore, a fixed N_{slice} was used to process the data at all growth stages. The ratio of plot length in the FOV to average inter-plant spacing was a reasonable reference. Hence, the number of volume slices was calculated as

$$N_{slice} = \left\lceil \frac{Plot\ Length\ in\ FOV}{Average\ Inter-plant\ Spacing} \right\rceil, \tag{1}$$

where ⌈ ⌉ denotes a ceiling function. Given that the imaged plot length was approximately 1.5 m and the average interplant spacing was 0.075 m, $N_{slice} = 20$ was used for our dataset. Figure 7 and Figure 8 demonstrate the PPH and PPW estimations visualized by the height and width of the bounding box for a short genotype and a tall genotype, respectively. The algorithm visits each point once for volume slicing and once for searching height/width in each volume slice. Weighted median computation is linear to N_{slice}. Since N_{slice} is significantly less than the point cloud size, the runtime is still linear to the point cloud size.
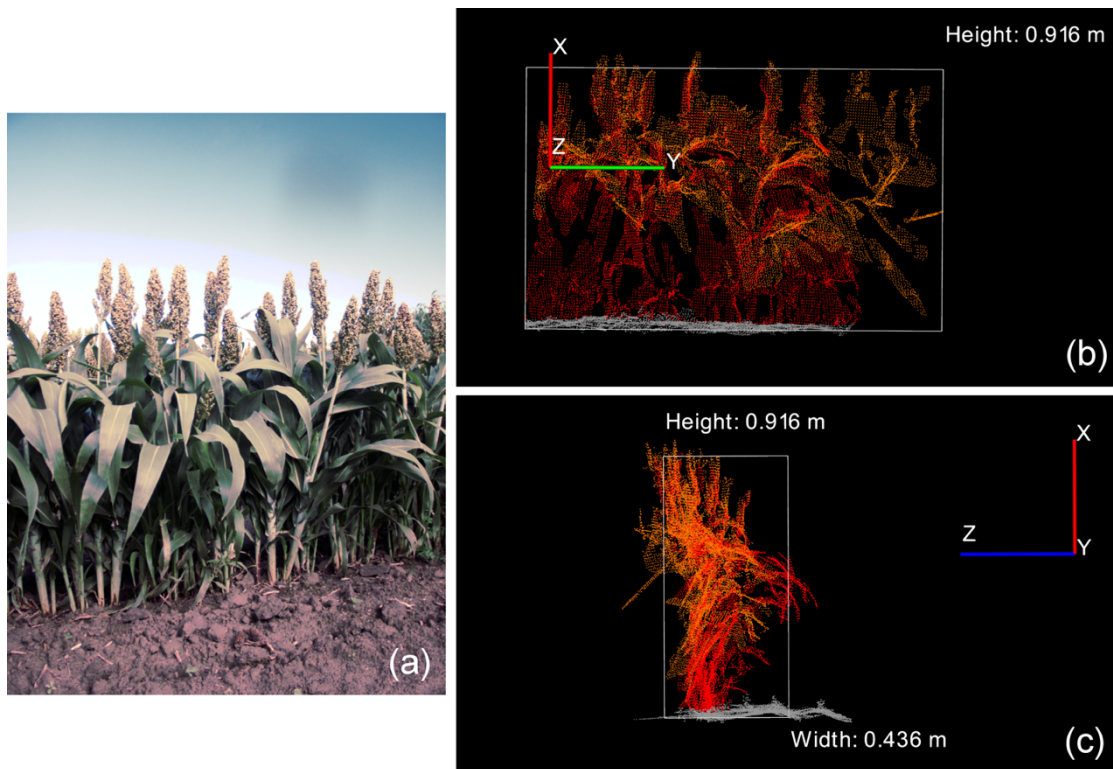
Figure 7: Plot-based plant height and plot-based plant width estimations for a short genotype. (a) The RGB image from the bottom stereo camera after histogram equalization for enhancement of the underexposed regions. (b) Front view of the 3D point cloud and bounding box. (c) Side view of the 3D point cloud and bounding box. Different colors of the point cloud represent the ground

surfaces (grey) and plant surfaces observed by the bottom (red) and middle (orange) stereo camera heads.
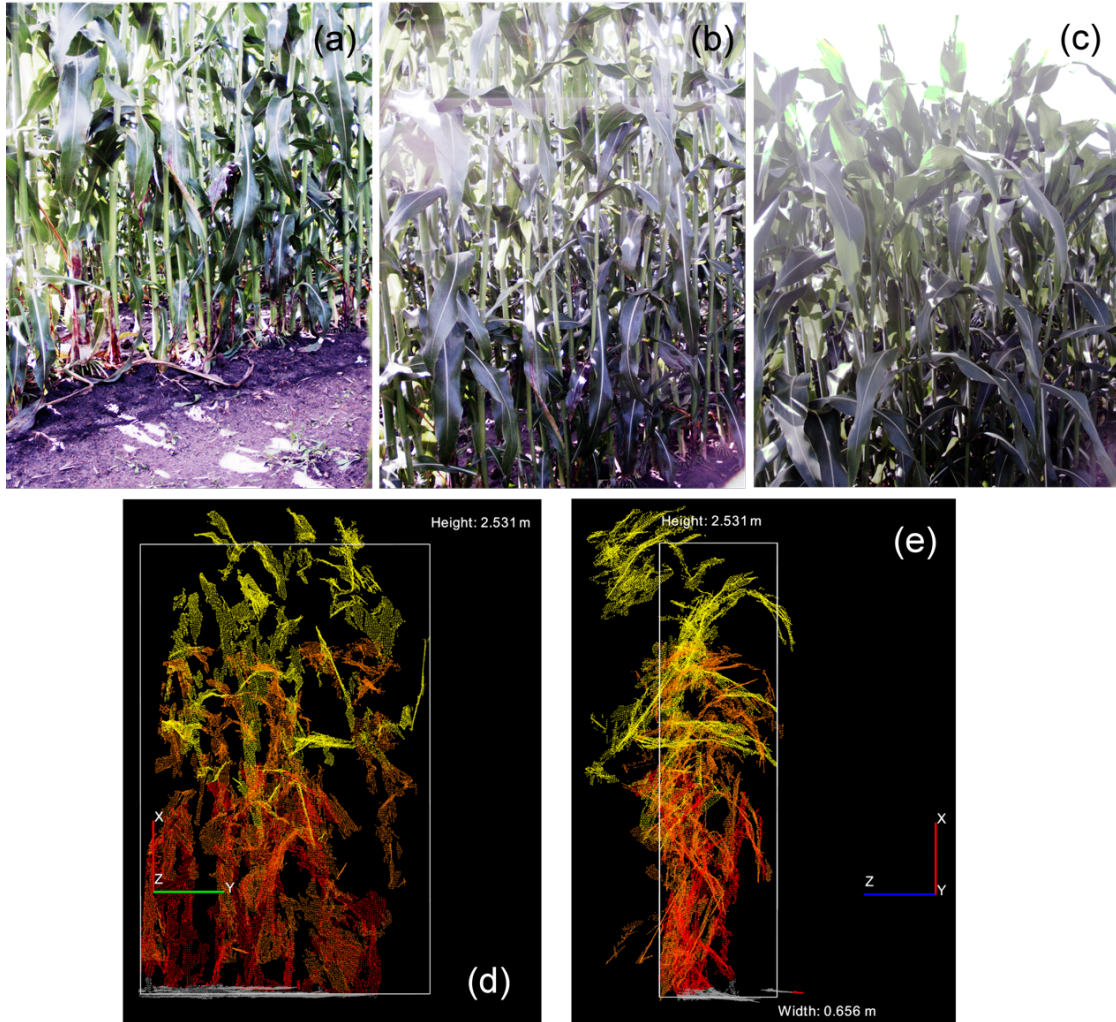

Figure 8: Plot-based plant height and plot-based plant width estimations for a tall genotype. (a), (b), and (c) are the enhanced RGB images from the bottom, middle, and top stereo camera heads, respectively. (d) Front view of the 3D point cloud and bounding box. (e) Side view of the 3D point cloud and bounding box. Different colors of the point cloud represent the ground surfaces (grey) and plant surfaces observed by the bottom (red), middle (orange), and top (yellow) stereo camera heads.

3.4. Convex Hull Volume and Plant Surface Area

Convex hull has been used to quantify canopy volume (Azzari, Goulden, & Rusu, 2013). Given a set of 3D points, the convex hull is the smallest convex set that contains all points. For any two points in a convex hull, the line segment connecting them must be in the convex hull. PCL relies on Qhull (Barber, Dobkin, & Huhdanpaa, 1996) to compute convex hull. Directly applying Qhull to the point cloud would not be accurate due to possible empty spaces in the plot. Therefore, the same slicing strategy was applied to identify slices that contained no plant. A convex hull was constructed for the points in each slice of the previously described bounding box. The volume ratio of the convex hull to its containing slice indicated the vegetation occupancy of the slice. If the ratio was lower than a threshold α, the convex hull was marked as invalid. We defined convex hull volume (CHV) as the sum of volumes from the valid convex hulls

$$CHV = \sum_{j=1}^{M} Volume_{valid\ chull}^{j}, \tag{2}$$

where M denotes the number of valid convex hulls. CHV requires two parameters, N_{slice} and minimum volume ratio α. The same N_{slice} in PPH and PPW estimation was used for CHV. Based on our dataset, we chose a heuristic value 0.3 for the minimum volume ratio α, which means that a valid volume slice should contain a set of points whose convex hull volume is larger than 30 percent of the slice volume. Figure 9 shows the valid convex hulls extracted from the short genotype and the tall genotype used in the previous section. The empty space in the point cloud was successfully detected with no convex hulls visualized in the corresponding volume slices. The empty space was caused by insufficient data points to form a plant shape.
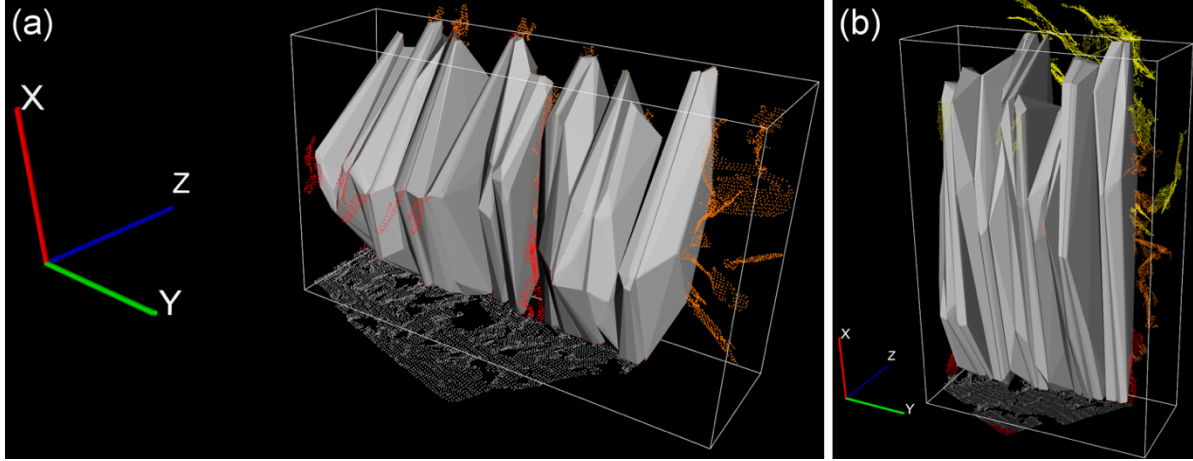
Figure 9: Valid convex hulls for (a) the short genotype in Figure 7 and (b) the tall genotype in Figure 8. Grey points represent ground surfaces, while other color points are either outside the bounding box or not enclosed by a valid convex hull.

To quantify plant surface area, the point cloud needs to be converted to a surface presentation. Triangle mesh is widely used for its simplicity and efficiency. PCL provides the GreedyProjectionTriangulation (Marton, Rusu, & Beetz, 2009) algorithm for fast surface reconstruction. MovingLeastSquares (Rusu & Cousins, 2011) was applied beforehand to smooth noisy surfaces, especially the overlapping surfaces captured by two adjacent stereo camera heads. This step improved the result of GreedyProjectionTriangulation. Given the three vertices (p_1, p_2, p_3) of a triangle, the triangle area (TA) was computed as

$$TA = \frac{|(p_2 - p_1) \times (p_3 - p_1)|}{2}. \tag{3}$$

The plant surface area (PSA) was approximated by the sum of areas of all triangles in the mesh

$$PSA = \sum_{l=1}^{O} TA_l, \tag{4}$$

where O denotes the number of triangles in the mesh. The PSA includes the surface area of leaves, stems, and panicles. Figure 10 shows the reconstructed triangle mesh for the short genotype and the tall genotype.
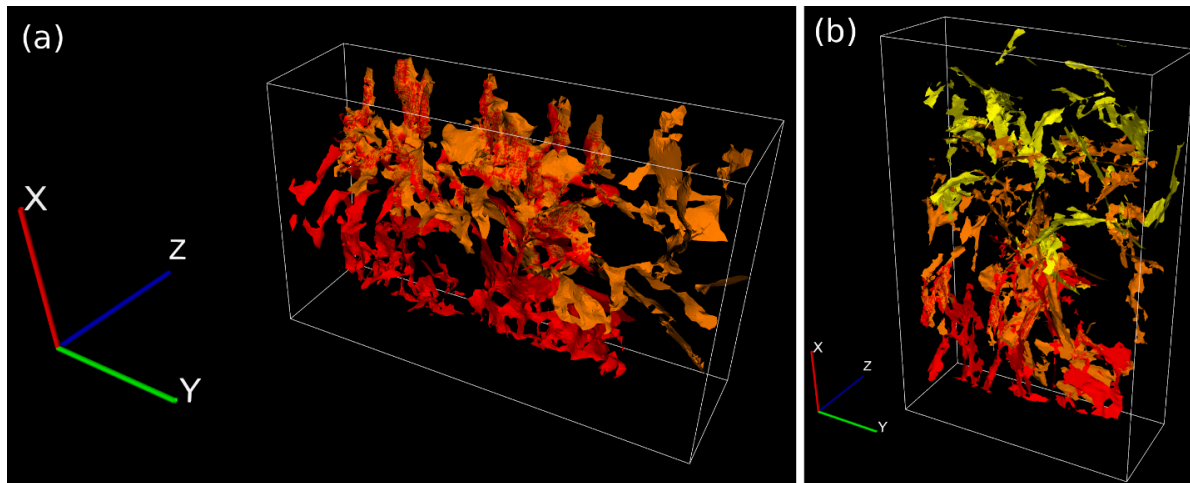

Figure 10: Triangle meshes for (a) the short genotype in Figure 7 and (b) the tall genotype in Figure 8. Surfaces outside the bounding box are discarded.

3.5. User-Interactive Stem Diameter Extraction

Extraction of sorghum stem diameter (SD) from the side-view images is a challenging task. First, sorghum has the ability to tiller. It is necessary to differentiate the stem of a plant from those of its tillers because tillers often have smaller SDs. Second, SDs should be measured at the same height above ground to provide comparable data. Third, the contour of the cross section of a sorghum stem (including the leaf sheaths) is best described as an ellipse. The difference between the maximum diameter and the minimum diameter could be large and the orientation of the plant should not be ignored. It is worth pointing out that the automated stem width estimation methods developed recently have shown promising accuracy (Baweja et al., 2018; Vijayarangan et al., 2018). However, currently they cannot differentiate the main plant from its tillers, which could

result in unwanted effects on the subsequent GWAS. Considering the above factors, we resorted to human decision to select which stem segment should be measured. A user interface was developed to facilitate the process in which a user is first shown three RGB images, as in Figure 6. Subsequently, the user must zoom in on a stem segment by clicking on it, and select four reference points to obtain the estimated SD (Fig. 11).

Figure 11: Four reference points on the stem edges chosen by the user.

The four points can be picked in any order as long as they form a convex quadrangle. Then, they are sorted such that $(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4)$ correspond to the top left corner, bottom left corner, top right corner, and bottom right corner in the image coordinate system (Fig. 12).
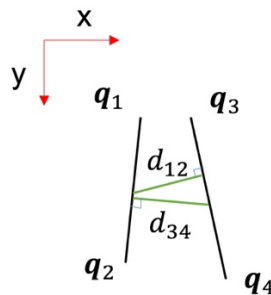

Figure 12: Sorted reference points and stem diameter estimation in the image coordinate system.

Let $\mathbf{q}_{12}$ denote the midpoint between $\mathbf{q}_1$ and $\mathbf{q}_2$, and $\mathbf{q}_{34}$ the midpoint between $\mathbf{q}_3$ and $\mathbf{q}_4$. The distance from $\mathbf{q}_{12}$ to the line passing through $\mathbf{q}_3$ and $\mathbf{q}_4$ was computed as

$$d_{12} = \frac{|(q_3-q_4)\times(q_4-q_{12})|}{|q_3-q_4|}. \tag{5}$$

Similarly, the distance from $\mathbf{q}_{34}$ to the line passing through $\mathbf{q}_1$ and $\mathbf{q}_2$ was

$$d_{34} = \frac{|(q_1-q_2)\times(q_2-q_{34})|}{|q_1-q_2|}. \tag{6}$$

The SD in the image coordinate system was estimated by

$$d = \frac{d_{12}+d_{34}}{2}. \tag{7}$$

The SD in metric unit (for instance, millimeter) was calculated as

$$D = \frac{bd}{d_{shared}}, \tag{8}$$

where b denotes the stereo camera baseline, and d_{shared} is the shared disparity of $(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4)$ (assuming the disparities are equal at the four points). The shared disparity d_{shared} was calculated by matching the image patch containing the stem segment in the reference (left) image of a stereo image pair to its correspondence in the target (right) image. Because the stem segment was relatively small compared to the full image size and there was not enough resolution to reconstruct the curved surface on the stem segment, it was reasonable to assume that $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$, and $\mathbf{q}_4$ were on the same plane parallel to the image plane. In other words, the depth values of all four reference points to the stereo camera were assumed to be equal. The stem segment patch matching score was evaluated using Normalized Cross-Correlation for its well-known robustness against radiometric differences in real-world images (Hirschmüller & Scharstein, 2007). The desired shared disparity was the one of the maximum matching score among all possible values.

3.6. Validation of the Image-Derived Features

A correlation analysis was performed between the in-field manual measurements and the image-derived features to identify highly correlated variables. This correlation would also provide valuable information to select image-derived features as proxies for biologically important traits that are targeted in genomic-based research. Repeatability of each image-derived feature was tested as well. Moreover, both 3DMST and SGBM were used to generate 3D point cloud data for feature extraction. The effect of the two stereo matching algorithms on the image-derived features were assessed.

The 18 plots (six genotypes and three replications) in 2016 were both imaged and manually measured twice during the growing season. The first measurement was 37 days after planting (DAP), while the second one was completed in two days, 98 and 99 DAP. We refer to the first date as the early stage and the second as the mature stage. Three sets of stereo images were collected at different positions along each plot. This was done by adding two additional imaging locations for each plot in the GPS-tagged imaging location map. One was 0.75 m before the center imaging location, and the other 0.75 m ahead. These images were used for the extraction of: i) PPH; ii) PPW; iii) CHV; and iv) PSA. The averaged results over the three positions for each plot were used for the correlation analysis.

The manually measured traits were collected on a single representative plant in each plot selected by trained plant breeders, and included the following: i) panicle apex height (PAH); ii) SD for every other internode; iii) leaf angle (LAN), leaf length (LL), leaf width (LW), leaf area (LAR), and leaf collar height (LCH) for each individual leaf; iv) leaf apex position (LAP) and leaf tip position (LTP) (or the 2D coordinates of the leaf apex and the leaf tip relative to the plant base

assuming the plant grows in a plane); and v) panicle length (PL) and panicle diameter (PD). Figure 13 illustrates the manually collected measurements on a sorghum plant, and Table 1 summarizes the available data for validating the automated image-derived features.
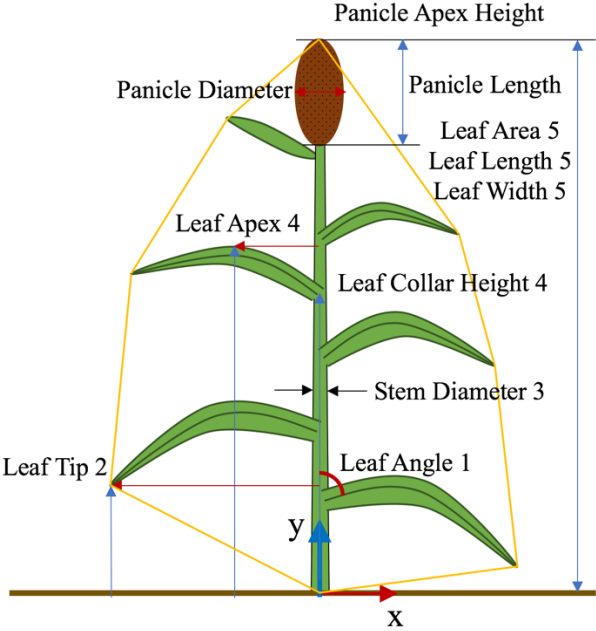

Figure 13: Illustration of the manually collected measurements on a sorghum plant. The plant is assumed to have a vertical stem. The 2D coordinates of each leaf collar, leaf apex, and leaf tip are measured with respect to the plant base.

Table 1. Summary of the available data for validating the automated image-derived features.

	Phenobot	In-field
Measurements per season	2	2
Measurements per plot	3	1
Total measurements	6	2

Features	PPH, PPW, CHV, PSA	PAH, SD, LAN, LL, LW, LAR, LCH, LAP, LTP, PL, PD

Note. CHV: convex hull volume; LAN: leaf angle; LAP: leaf apex position; LAR: leaf area; LCH: leaf collar height; LL: leaf length; LTP: leaf tip position; LW: leaf width; PAH: panicle apex height; PD: panicle diameter; PL: panicle length; PPH: plot-based plant height; PPW: plot-based plant width; PSA: plant surface area; SD: stem diameter.

The height and width measurements were done in situ by using a measuring stick, and leaf angle using a protractor. Then the plants were taken to a lab at the farm. Each leaf was cut off from the plant at the collar, and the LL was measured from leaf base to leaf tip with a string closely following the midrib. Leaf area was scanned by a leaf area meter. After removal of the leaf sheaths from the stem, the length of each internode and the SD of every other internode were obtained using a tape measure and a caliper, respectively. Stem surface area was approximated by the sum of the cylinder surface area of each internode. Similarly, panicle surface area was calculated as the surface area of the cylinder parameterized by the PL and diameter at its widest point. Hence, the manually measured plant surface area equaled the sum of surface area of leaves, stem, and panicle. Given the 2D coordinates of the leaf apexes and PAH (assuming the stem was vertical) mentioned above, the 2D bounding shape of a representative plant was approximated as the 2D convex hull of those coordinates (Fig. 13) using the Qhull algorithm. The 3D CHV was calculated as the product of the 2D convex hull area and the average interplant spacing. Since the length of a plot in the image was approximately constant, the image-derived CHV was expected to correlate well with the convex hull volume of the presentative plant.

The user-interactive SD was assessed using an earlier dataset. The SDs of a subset of 20 genotypes were measured using a caliper in two field locations during the summer in 2014 at two different time points: 63-75 DAP and 83-87 DAP. A total of 80 in-field measurements were acquired. The measured stalk segments were marked with red ribbons so they could be identified in the images. Because the cross section of a sorghum stem (with leaf sheaths) is not circular, the in-field caliper measurements were performed along the row direction as the camera would capture. The image-derived SD data were obtained four times from each stem, with the user-interactive approach described above.

To evaluate repeatability of the image-derived features, standard deviation (*STD*) and coefficient of variation (*CV*) are reported based on the deviation between each image-derived feature measurement and the average value for its plot, since the true value is unknown. The *STD* is the standard deviation of all measurement deviations. The *CV* is calculated by dividing the *STD* of the deviations by the mean for each plot and then taking the mean of these values over all plots (Underwood et al., 2017):

$$CV = mean_i \left(\frac{STD_j(dv_{ji})}{mean_j(dv_{ji})} \right) \times 100\%, \tag{9}$$

where the subscript *i* is the plot ID, the subscript *j* is the measurement ID, dv_{ji} is the measurement deviation *j* of an image-derived feature (e.g., PPH, PPW, CHV, PSA, or SD) for plot *i*.

4. RESULTS

This section includes correlation and repeatability analyses of the image-derived features. Additionally, a detailed time analysis of the data processing pipeline is presented, followed by a time comparison between the human labor and the robotic solution.

4.1. Plot-based Plant Height and Plot-based Plant Width

Before flowering, measuring plant height is an ambiguous task due to the lack of a reliable reference as the top of a plant. Thus, both the maximum canopy height and leaf collar height of the last fully expanded leaf were analyzed for the early stage. The correlation between the image-derived PPH and the manually collected plant height was moderate at the early stage (Fig. 14 left and middle). In contrast, the PPH and the panicle apex height had a very strong linear relationship at the mature stage with either stereo matching algorithm ($R_{3DMST} = 0.97$, $R_{SGBM} = 0.98$) (Fig. 14 right). Note that the datasets of both growth stages were not combined for further assessment because different traits were manually measured. To assess repeatability, the PPH data from both stages were combined. The *STD* and *CV* were 0.05 m and 4.7%, respectively when using 3DMST, while SGBM resulted in similar repeatability with *STD* = 0.04 m and *CV* = 3.8%.
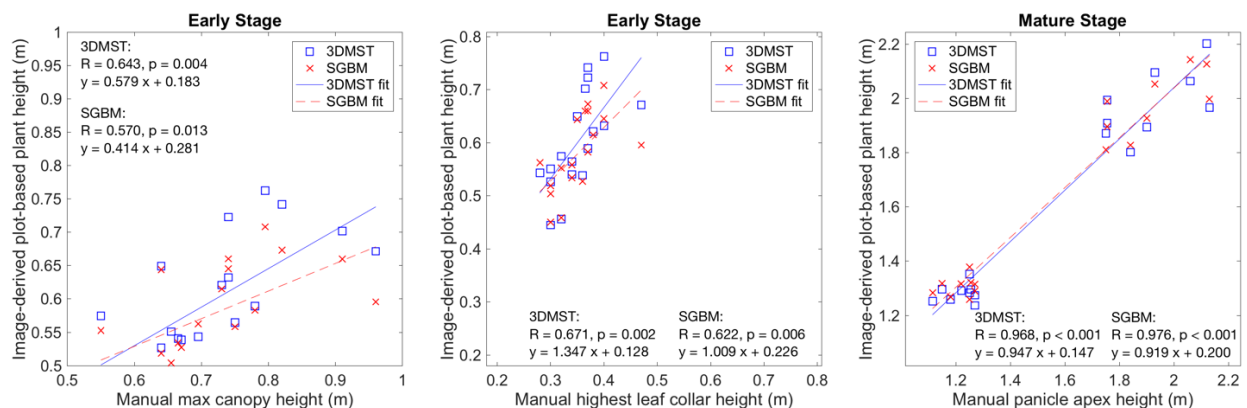

Figure 14: Correlation analysis of the image-derived plot-based plant height (PPH) with two stereo matching algorithms (3DMST and SGBM). Left: the manually measured max canopy height vs. the PPH at the early stage. Middle: the manually measured highest leaf collar height

vs. the PPH at the early stage. Right: the manually measured panicle apex height vs the PPH at

the mature stage. There are 18 samples including six genotypes and three replications.

At the early stage, the image-derived PPW was not correlated with the average leaf length with

the p-values larger than 0.05 (Fig. 15 left). At the mature stage, the PPW was moderately

correlated with the average leaf length (Fig. 15 middle). However, if the datasets of both growth

stages were combined, the PPW achieved a very strong correlation with the average leaf length

(Fig. 15 right). The *STD* and *CV* were 0.03 m and 8.6% for 3DMST, and 0.04 m and 9.8% for

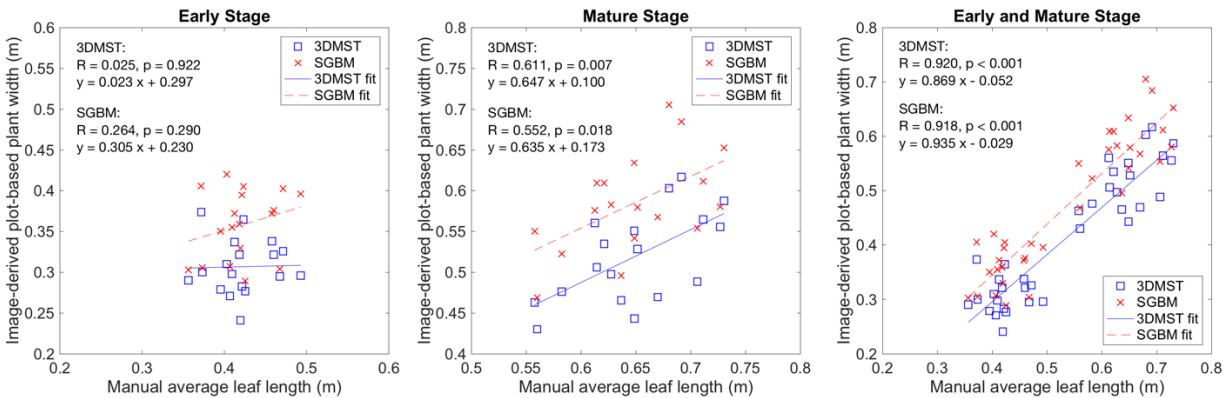SGBM. Both 3DMST and SGBM performed closely on this feature.

Figure 15: Correlation analysis of the image-derived plot-based plant width (PPW) with two

stereo matching algorithms (3DMST and SGBM). The PPW is compared with the manually

measured average leaf length. Left: the early stage. Middle: the mature stage. Right: the early

and mature stages combined. There are 18 samples including six genotypes and three

replications.

4.2. Convex Hull Volume and Plant Surface Area

At the early stage, there was no correlation between the image-derived CHV and the manually

obtained convex hull volume of the single representative plant (Fig. 16 top left) at the 5%

significance level. However, the correlation was moderate at the mature stage (Fig. 16 top

middle), and became very strong when both growth stages were combined (Fig. 16 top right).

The *STD*s were both 0.03 m^3 for the two methods, but the *CV*s were quite different with 17.8%

for 3DMST and 10.7% for SGBM. The reason was that using SGBM resulted in much higher

CHVs than using 3DMST did, which caused much lower *CV*s for SGBM with similar within-

plot *STD*s to 3DMST. The higher CHV using SGBM was probably due to its noisier 3D

reconstruction which expanded the envelop of the point cloud. Alternatively, the product of the

PPH and the PPW achieved a slightly higher correlation with the manually obtained 2D convex

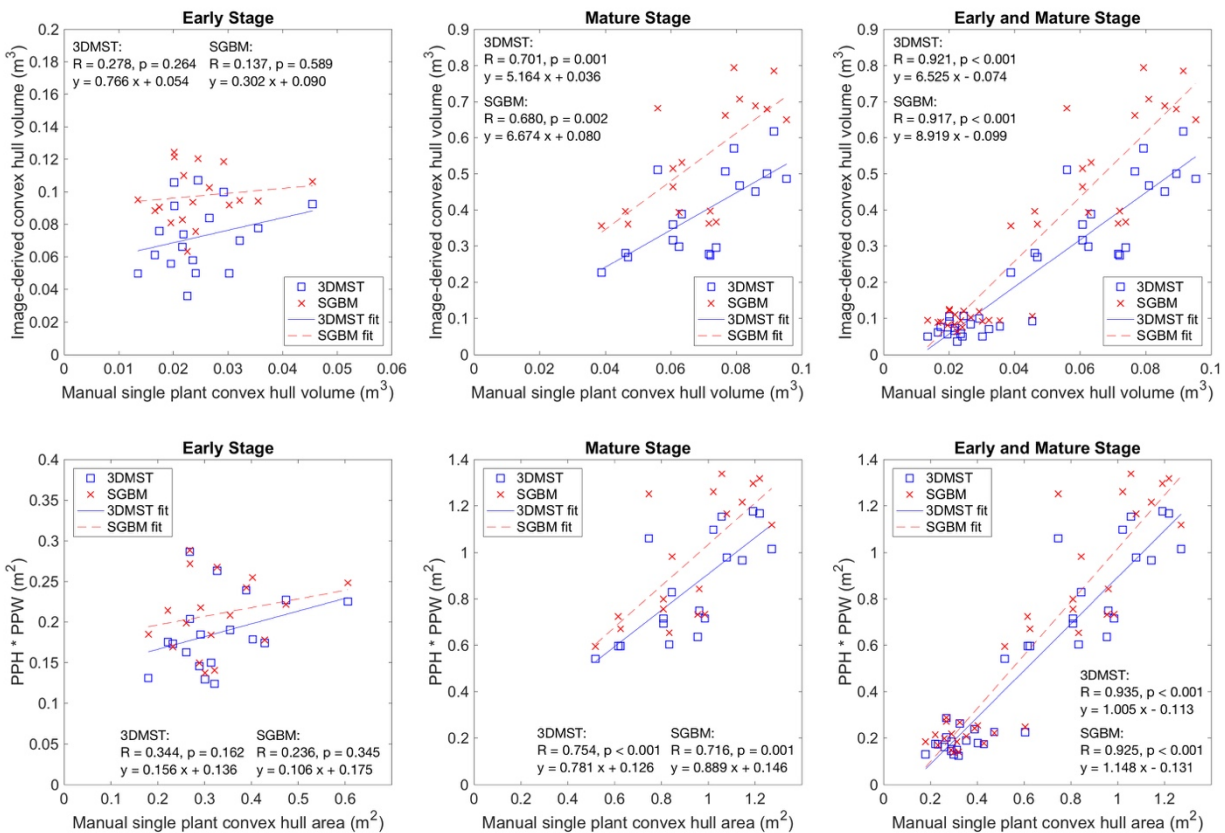hull area than the CHV did with the manually obtained convex hull volume (Fig. 16 bottom

row).

Figure 16: Correlation analyses of the image-derived convex hull volume (CHV) and the product of the plot-based plant height (PPH) and the plot-based plant width (PPW) with two stereo matching algorithms (3DMST and SGBM). Top row: the manually measured convex hull volume of the single representative plant vs. the image-derived CHV. Bottom row: the manually measured 2D convex hull area of the single representative plant vs. the product of the PPH and the PPW. There are 18 samples including six genotypes and three replications.

At the early stage, the image-derived PSA was not correlated with the manually measured surface area of the single representative plant. However, at the mature stage, a very strong linear relationship ($R_{3DMST} = 0.82$) was found when 3DMST was used for 3D reconstruction, much higher than that obtained with SGBM ($R_{SGBM} = 0.66$) (Fig. 17 top row). Regarding repeatability, the *STD* and *CV* were 0.08 m^2 and 8.7% for 3DMST while larger values were obtained for SGBM (*STD* = 0.11 m^2 and *CV* = 9.1%). The results suggested that 3DMST improved the accuracy of PSA estimation. As illustrated in Figure 4, large variations were observed in the surface normal maps generated by SGBM, which was the most likely reason behind the overall higher PSA and lower repeatability associated with SGBM. When both growth stages were combined, the image-derived PSA had a very strong correlation with the manual measured one for both 3DMST and SGBM. Additionally, there was no correlation between the convex hull surface area and the single plant surface area at the 5% significance level except when both growth stages were combined (Fig. 17 bottom row).
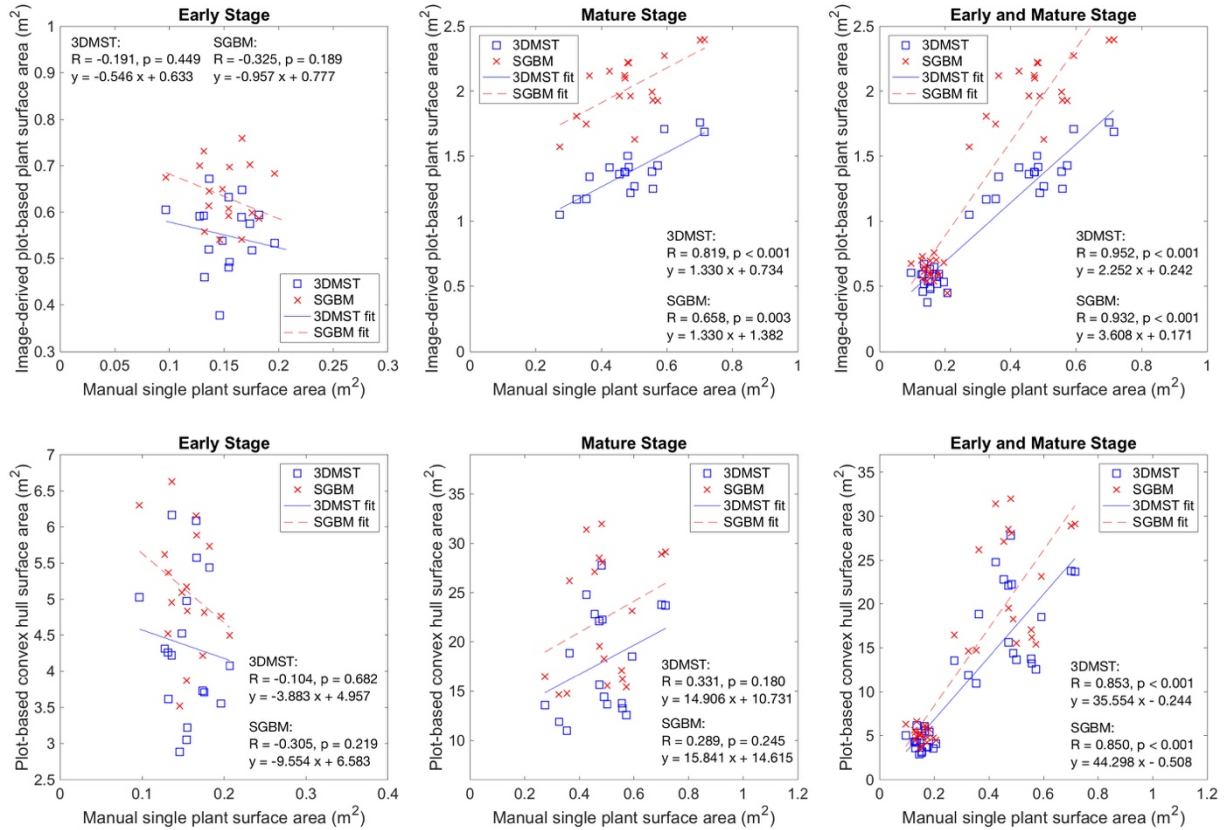
Figure 17: Correlation analyses of the image-derived plant surface area (PSA) and the convex hull surface area with two stereo matching algorithms (3DMST and SGBM). Top row: the manually measured surface area of the single representative plant vs. the image-derived PSA. Bottom row: the manually measured surface area of the single representative plant vs. the image-derived convex hull surface area. There are 18 samples including six genotypes and three replications.

4.3. Stem Diameter

The image-derived SD had a very strong correlation with the manually measured SD ($R = 0.96$) (Fig. 18), even though the image-derived approach produced larger values than the caliper measurements, as demonstrated by the intercept of the fitted line (2.8 mm). One possible explanation is that the caliper could be pressed against the soft leaf sheaths around the stem,

resulting in smaller values. The user-interactive approach was also highly repeatable with a

STD of 0.70 mm and a *CV* of 3.8%. In comparison with the current automated methods, the

user-interactive method demonstrated higher accuracy. The StalkNet achieved a mean absolute

error (MAE) of 2.76 mm (Baweja et al., 2018), while the user-interactive method achieved a

MAE of 1.64 mm. If the range of the ground truth SDs was considered, the MAE of the

StalkNet equaled 29.05% of their ground truth SD range (minimum = 8.00 mm, maximum =

17.50 mm). However, the MAE of the user-interactive method equaled 7.07% of our ground

truth SD range (minimum = 8.47 mm, maximum = 31.67 mm). Vijayarangan et al. (2018)

reported a normalized root mean squared error of 25.15% for their automated SD estimation

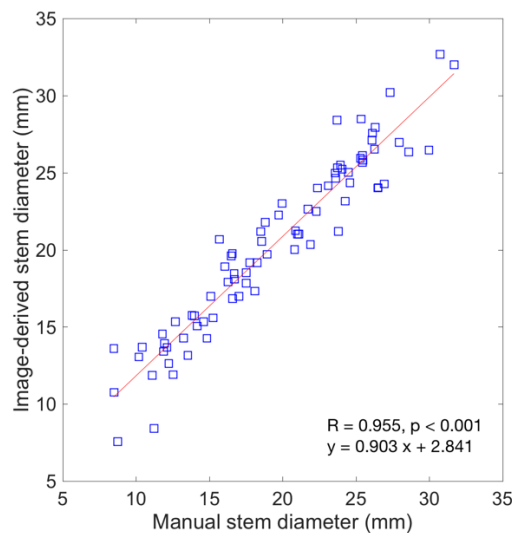method, while the same measure for the user-interactive method was 8.65%.

Figure 18: Correlation analysis of the user-interactive image-derived stem diameter (SD). There

are 80 samples (20 genotypes × 2 locations × 2 time points). The image-derived SD values

showed a strong positive correlation with the manually measured SD.

4.4. Time Analysis

The stereo image data were processed on a desktop workstation with a 3.5 GHz Xeon HexaCore CPU, 16 GB RAM, and a NVIDIA GTX Titan X GPU (Maxwell). The stereo images were downscaled to the resolution of 812×612. The average stereo matching time for 3DMST was 30 s per stereo pair, while it was 0.4 s for SGBM. 3DMST was 75 times slower than SGBM in our application. The time analysis of the data processing pipeline is presented in Table 2. Due to the long runtime of 3DMST, the processing pipeline was dominated by the stereo matching stage, which was not the case for SGBM occupying only 17% of the total processing time. In the point cloud processing (Stage 2-5 in Table 2), the most time-consuming stage was the PSA extraction mostly due to the expensive GreedyProjectionTriangulation. The total processing time increased approximately linear to plant height as expected.

Table 2. Time analysis of the data processing pipeline.

Processing Stage \ Plant Height	1.1 m	1.8 m	2.7 m
1. Stereo matching (3DMST/SGBM)	30.214/0.407	60.405/0.815	90.643/1.225
2. Preprocessing	0.689	1.055	1.333
3. PPH and PPW	0.002	0.004	0.007
4. CHV	0.021	0.036	0.050
5. PSA	1.168	3.001	4.062
Stage 2-5	1.880	4.096	5.452
Total	32.094/2.287	64.501/4.911	96.095/6.675

Regarding human labor, there were eight to ten people taking in-field manual measurements in the 2016 validation trial. At the early stage, it took a full day to measure the traits for the 18 plots, and the estimated man-hours were approximately 32. However, at the mature stage, the man-hours doubled to 64 since the plants grew taller with more leaves and internodes to measure. Whereas, the Phenobot took approximately 3 min to collect the stereo images for the 18 plots, which was determined by the number of plots, the plot length, and the maximum travel speed regardless of growth stage. The automated data processing at the mature stage took 20 min and 83 s using 3DMST and SGBM, respectively. Using the user-interactive SD extraction software, it typically took 10 s for human to measure a stem segment. Although the SD extraction was not fully automated, the robotic phenotyping system still demonstrated its superior efficiency over the current manual practice.

5. DISCUSSION

In this section, the data acquisition system, stereo matching algorithms, and image-derived features are discussed based on our experimental results. Limitations and potential solutions are also suggested for future improvement.

5.1. Phenobot 1.0

The Phenobot 1.0 has been successfully deployed for two growing seasons and collected over 100,000 stereo images for various genetic lines. During this process, technical challenges had to be overcome and methods adapted to the biology of the target organism. Biomass sorghum is a

tall crop with dense canopies, making it one of the most difficult plant species for ground-based high-throughput phenotyping. At the mature stage, plant height ranges from 0.5 m to more than 3 m, with leaves that grow vertically and expand horizontally. Considering that these long leaves would likely block the camera view if the commercial row spacing (0.76 m) was implemented, we compromised the standard row spacing and expanded it to 1.5 m during the system testing phase. However, as the leaves were expanding, it was evident that they were still blocking the view of the mid-level and top-level cameras, which led to the selection of a final 2.2 m row spacing with which only occasional view blocking of the top-level cameras occurred. One alternative solution that could potentially reduce row spacing would be to image multiple locations per plot and to place additional sets of stereo camera heads vertically with varying vertical tilting angles.

The data collection time was limited between 10 AM and 4 PM to avoid low solar elevation angles, which could cause strong backlighting conditions, especially if crop rows were planted north to south. During normal operation time, the varying field lighting conditions did not pose a challenge for stereo reconstruction because stereo matching relied on image texture or local intensity variation, which was preserved as long as the amount of light received by the imaging sensor was within its dynamic range. However, we recognize that it would be beneficial to add the active lighting (Mueller-Sim et al., 2017; Pothen & Nuske, 2016) for the stereo camera heads to extend the limitation on operation time. Moreover, extremely tall genotypes could substantially block the sunlight, and the image noise level increases in such low-light conditions, which could greatly affect the stereo reconstruction. Another potential improvement to the in-field active lighting for stereo cameras would be to combine the high-powered flash, the structured light pattern (Nguyen et al., 2016), and polarizing filters. The high-powered flash with

short camera exposure times can reduce the effect of varying ambient light. However, sorghum plant surfaces tend to be reflective. Polarizing filters can be added in front of both the flash and the camera lens to suppress glare and improve image quality.

5.2. Stereo Matching

There has been a concern about the performance of passive stereo vision under field lighting conditions (Kazmi, Foix, Alenyà, & Andersen, 2014; Li, Zhang, & Huang, 2014). However, our results demonstrate that the advances in stereo matching have allowed for the reconstruction of high quality disparity maps for dense canopy crops in the field. As for the performance difference between the state-of-the-art 3DMST and the widely used SGBM, although a quantitative comparison is not provided, our qualitative comparison of disparity maps and surface normal maps, and the correlation analysis for plant surface area validate the superior surface reconstruction quality of 3DMST over that of SGBM. First, 3DMST clearly produces more accurate contours of leaf, stalk, and panicle. On the contrary, SGBM slightly extends foreground disparity on the edges into the background, which is often referred to as border bleeding effects. This is caused by the lack of adaptive support weight in SGBM. On the other hand, the geodesic type of adaptive support weight utilized by 3DMST accurately captures the depth discontinuity. Such property could be more valuable for applications where accurate detection and measurement of stem diameter are needed. Second, a much smoother normal estimation is achieved on plant surface when 3DMST is used instead of SGBM. The non-local support region in 3DMST can adapt the effective area of support region to image content. As the effective support region becomes larger, the smoothness strength is implicitly increased as well. Hence, 3DMST easily handles large canopies and thin stems in nearly homogenous color. In addition, 3DMST jointly searches the disparity and surface normal (3D label), which are in turn

used to calculate a more accurate support region mapping between the left and the right images. For instance, a square support region in the left image is scaled or sheared in the right image, depending on the surface normal direction (Heise, Klose, Jensen, & Knoll, 2013). Without this more flexible model, disparity on slanted leaf surfaces would show stair-casing effects. Even though SGBM uses scanline optimization to enforce smoothness along multiple directions and applies quadratic interpolation around the optimal discrete disparity level to obtain sub-pixel accuracy, it is difficult to achieve the same smoothness level of 3DMST, even by tuning the parameters. From all features investigated in this study, plant surface area was the parameter mostly affected by the selection of the 3D reconstruction algorithm. The edge-preserving ability and surface smoothness caused a considerable impact on the surface area calculated from the triangle mesh. This was revealed by the correlation comparison demonstrating that the image-derived plant surface area using SGBM was approximately $0.67\ m^2$ larger than that of 3DMST (Fig. 17 top middle).

Regarding runtime performance, SGBM has a clear advantage over 3DMST due to its high computational efficiency. The implementation of SGBM takes full advantage of instruction level parallelism on modern CPUs, also known as single instruction, multiple data. Therefore, if the goal is only to estimate features not related to surface such as plot height, plot width, and convex hull volume, SGBM should be the preferred option. There is still potential to cut down the runtime of 3DMST by reducing the number of iterations and using a simple matching cost such as Census. However, the effect of these two factors on the reconstruction quality is unknown and should be further investigated.

5.3. Image-Derived Plant Architectural Features

The correlation analyses demonstrate that the image-derived features investigated in this study can capture several important plant architectural traits of densely populated sorghum varieties grown under field conditions.

Plant height is probably the most important characteristic due to its significant effect on biomass yield. Although the image-derived PPH was obtained on a per-plot basis, it was highly correlated with the manually measured plant height. The PPW turned out to be useful for quantifying leaf length variation, which could affect the ability of sorghum to absorb light energy. Considering that the PPW could also be determined by leaf angle, if the leaf length is constant, a large angle would result in a large PPW.

The convex hull volume combines the effects of plant height, leaf position, and leaf shape. For instance, some sorghum varieties have a long exsertion (distance from flag leaf to panicle) while others have no exsertion at all, and thus, the panicle is in close proximity to the canopy. Given the same plant height, the image-derived CHV could capture such a variation in plant architecture.

The plant surface area can be affected by many factors such as number of leaves, internode length, leaf area, plant height, stem diameter, and panicle size. Even though the cameras observed the plot, the extracted PSA correlated well with the counterpart of the representative individual.

Stem diameter is one of the most difficult traits to characterize using high-throughput phenotyping, but also among the most important plant architecture traits since it contributes greatly to biomass yield. The variation in stem diameter was evident in our dataset (from 7 mm to 31 mm), and the user-interactive approach produced accurate measurements. It is important to

emphasize that the image-derived SD included the thickness contributed by the leaf sheath around it, and thus it could be used as proxy of the true stem diameter, which should be measured after the removal of the leaf sheath. Although the user-interactive approach cannot match the speed of the automated method (Baweja et al., 2018), it is still much faster and less laborious than in-field manual measuring.

The proposed features have been partially validated by our recently published GWAS performed using the image-derived PPH and SD after maturity (Salas Fernandez, Bao, Tang, & Schnable, 2017). Genomic regions controlling the two traits estimated using our approach co-localized with genetic markers previously identified associated with variation in plant height and stem diameter using manually collected data (Zhao et al., 2016). Additionally, this also eases the concern that our increased row spacing would affect the GWAS results, since the data used to validate our GWAS was collected from materials planted at the commercially used row spacing of 0.76 m. It is worth pointing out that the discriminant ability of these image-derived features depends on the plant growth stage. At early stages, the variation among crop varieties could be too low to be captured by our approach. As the growing season progresses and differences in architecture and development become evident, the image-derived features could reach the maximum potential to quantify these traits and characterize genetic variation. Our ongoing work includes the discovery of genetic regions that control the variations observed in image-derived PPW, CHV, and PSA. Additionally, the PPW, CHV, and PSA values at different canopy levels can also be quantified, which could potentially provide novel information about plant architecture for gene discovery studies.

Our feature extraction pipeline is fully automated except for stem diameter that requires human intervention to select stems and indicate stem boundaries. The automated StalkNet (Baweja et

al., 2018) has already shown promising accuracy for stem instance segmentation. We believe that the state-of-the-art Mask R-CNN (He, Gkioxari, Dollár, & Girshick, 2017) has great potential for superior stem segmentation. With the large number of side-view images collected in the field, a large amount of training data can be generated to explore the performance of Mask R-CNN for instance segmentation of stem and beyond. For instance, panicle instance segmentation can reveal flowering time and facilitate grain yield prediction.

6. CONCLUSIONS

This paper presents a field-based robotic plant phenotyping system that utilized 3D machine vision to automate plant architectural trait characterization for densely populated biomass sorghum plants, with a wide range of plant height. The data acquisition system was developed based on a utility tractor retrofitted with an auto-steer system and multi-level side-viewing stereo camera heads. An automated feature extraction pipeline was developed to quantify plot-based plant height, plot-based plant width, convex hull volume, and plant surface area. Stem diameter was extracted in a user-interactive approach. The image-derived features were highly correlated with in-field manual measurements with high repeatability. Our study also demonstrated the suitability of conventional passive stereo vision for 3D plant phenotyping under field lighting conditions due to the recent advances in stereo correspondence search algorithms. The proposed system demonstrates a great potential for large-scale field-based high-throughput plant phenotyping for bioenergy crops. Our future work will focus on building an autonomous ground robot, an imaging system optimized for commercial row spacing, and the automatic detection and characterization of individual plant organs such as stem, leaf, and panicle.

ACKNOWLEDGEMENTS

REFERENCES

Andrade-Sanchez, P., Gore, M. A., Heun, J. T., Thorp, K. R., Carmo-Silva, A. E., French, A. N., … White, J. W. (2014). Development and evaluation of a field-based high-throughput phenotyping platform. *Functional Plant Biology*, *41*(1), 68–79.

Azzari, G., Goulden, M. L., & Rusu, R. B. (2013). Rapid characterization of vegetation structure with a Microsoft Kinect sensor. *Sensors*, *13*(2), 2384–2398.

Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)*, *22*(4), 469–483.

Bargoti, S., Underwood, J. P., Nieto, J. I., & Sukkarieh, S. (2015). A pipeline for trunk localisation using LiDAR in trellis structured orchards. *Field and Service Robotics. Springer Tracts in Advanced Robotics*, *105*, 455–468.

Baweja, H. S., Parhar, T., Mirbod, O., & Nuske, S. (2018). StalkNet: A Deep Learning Pipeline

for High-Throughput Measurement of Plant Stalk Count and Stalk Width. *Field and Service Robotics. Springer Proceedings in Advanced Robotics*, *5*, 271–284.

Bleyer, M., Rhemann, C., & Rother, C. (2011). PatchMatch Stereo-Stereo Matching with Slanted Support Windows. In *Proceedings of the British Machine Vision Conference* (Vol. 11, pp. 1–11). Dundee, Scotland: BMVC Press.

Bradski, G., & Kaehler, A. (2000). OpenCV. *Dr. Dobb's Journal of Software Tools*, *3*.

Chaivivatrakul, S., Tang, L., Dailey, M. N., & Nakarmi, A. D. (2014). Automatic morphological trait characterization for corn plants via 3D holographic reconstruction. *Computers and Electronics in Agriculture*, *109*, 109–123.

Chapman, S. C., Merz, T., Chan, A., Jackway, P., Hrabar, S., Dreccer, M. F., … Jimenez-Berni, J. (2014). Pheno-copter: a low-altitude, autonomous remote-sensing robotic helicopter for high-throughput field-based phenotyping. *Agronomy*, *4*(2), 279–301.

Fischler, M. A., & Bolles, R. C. (1987). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. In *Readings in Computer Vision* (pp. 726–740). New York: Elsevier.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Computer Vision (ICCV), 2017 IEEE International Conference on* (pp. 2980–2988). Venice, Italy: Piscataway, NJ: IEEE.

Heise, P., Klose, S., Jensen, B., & Knoll, A. (2013). Pm-huber: Patchmatch with huber regularization for stereo matching. In *Computer Vision (ICCV), 2013 IEEE International Conference on* (pp. 2360–2367). Sydney, Australia: Piscataway, NJ: IEEE.

Hirschmüller, H. (2008). Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2), 328–341.

Hirschmüller, H., & Scharstein, D. (2007). Evaluation of cost functions for stereo matching. In *Computer Vision and Pattern Recognition (CVPR), 2007 IEEE Conference on* (pp. 1–8). Minneapolis, MN: Piscataway, NJ: IEEE.

Hirschmüller, H., & Scharstein, D. (2009). Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(9), 1582–1599.

Hosni, A., Bleyer, M., Gelautz, M., & Rhemann, C. (2009). Local stereo matching using geodesic support weights. In *Image Processing (ICIP), 2009 IEEE International Conference on* (pp. 2093–2096). Cairo, Egypt: Piscataway, NJ: IEEE.

Hosni, A., Rhemann, C., Bleyer, M., Rother, C., & Gelautz, M. (2013). Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(2), 504–511.

Jiang, Y., Li, C., & Paterson, A. H. (2016). High throughput phenotyping of cotton plant height using depth images under field conditions. *Computers and Electronics in Agriculture*, *130*, 57–68.

Jiang, Y., Li, C., Robertson, J. S., Sun, S., Xu, R., & Paterson, A. H. (2018). GPhenoVision: A Ground Mobile System with Multi-modal Imaging for Field-based High Throughput Phenotyping of Cotton. *Scientific Reports*, *8*(1), 1213.

Kazmi, W., Foix, S., Alenyà, G., & Andersen, H. J. (2014). Indoor and outdoor depth imaging of

leaves with time-of-flight and stereo vision sensors: Analysis and comparison. *ISPRS Journal of Photogrammetry and Remote Sensing*, *88*, 128–146.

Li, J., & Tang, L. (2017). Developing a low-cost 3D plant morphological traits characterization system. *Computers and Electronics in Agriculture*, *143*, 1–13.

Li, L., Yu, X., Zhang, S., Zhao, X., & Zhang, L. (2017). 3D cost aggregation with multiple minimum spanning trees for stereo matching. *Applied Optics*, *56*(12), 3411–3420.

Li, L., Zhang, Q., & Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors*, *14*(11), 20078–20111.

Lou, L., Liu, Y., Han, J., & Doonan, J. H. (2014). Accurate multi-view stereo 3D reconstruction for cost-effective plant phenotyping. In *Image Analysis and Recognition, 2014 International Conference on* (pp. 349–356). Vilamoura, Algarve, Portugal: Cham: Springer.

Marton, Z. C., Rusu, R. B., & Beetz, M. (2009). On fast surface reconstruction methods for large and noisy point clouds. In *Robotics and Automation (ICRA), 2009 IEEE International Conference on* (pp. 3218–3223). Kobe, Japan: Piscataway, NJ: IEEE.

Mueller-Sim, T., Jenkins, M., Abel, J., & Kantor, G. (2017). The Robotanist: a ground-based agricultural robot for high-throughput crop phenotyping. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on* (pp. 3634–3639). Marina Bay Sands, Singapore: Piscataway, NJ: IEEE.

Multani, D. S., Briggs, S. P., Chamberlin, M. A., Blakeslee, J. J., Murphy, A. S., & Johal, G. S. (2003). Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. *Sci*, *302*(5642), 81–84.

Nguyen, T. T., Slaughter, D. C., Maloof, J. N., & Sinha, N. (2016). Plant phenotyping using multi-view stereo vision with structured lights. In *Proc. SPIE 9866, Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping* (Vol. 9866, p. 986608). Baltimore, Maryland: Bellingham, WA: The International Society for Optics and Photonics.

Olson, S. N., Ritter, K., Rooney, W., Kemanian, A., McCarl, B. A., Zhang, Y., … Mullet, J. (2012). High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops. *Biofuels, Bioproducts and Biorefining*, *6*(6), 640–655.

Olsson, C., Ulén, J., & Boykov, Y. (2013). In defense of 3d-label stereo. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on* (pp. 1730–1737). Portland, OR: Piscataway, NJ: IEEE.

Pothen, Z. S., & Nuske, S. (2016). Texture-based fruit detection via images using the smooth patterns on the fruit. In *Robotics and Automation (ICRA), 2016 IEEE International Conference on* (pp. 5171–5176). Stockholm, Sweden: Piscataway, NJ: IEEE.

Ranftl, R., Bredies, K., & Pock, T. (2014). Non-local total generalized variation for optical flow estimation. In *European Conference on Computer Vision* (pp. 439–454). Zürich, Switzerland: Cham: Springer.

Rooney, W. L., Blumenthal, J., Bean, B., & Mullet, J. E. (2007). Designing sorghum as a dedicated bioenergy feedstock. *Biofuels, Bioproducts and Biorefining*, *1*(2), 147–157.

Ruckelshausen, A., Biber, P., Dorna, M., Gremmes, H., Klose, R., Linz, A., … others. (2009). BoniRob--an autonomous field robot platform for individual plant phenotyping. *Precision Agriculture*, *9*(841), 1.

Rusu, R. B. (2010). Semantic 3D object maps for everyday manipulation in human living environments. *KI-Künstliche Intelligenz*, *24*(4), 345–348.

Rusu, R. B., & Cousins, S. (2011). 3d is here: Point cloud library (pcl). In *Robotics and Automation (ICRA), 2011 IEEE International Conference on* (pp. 1–4). Shanghai, China: Piscataway, NJ: IEEE.

Rusu, R. B., Marton, Z. C., Blodow, N., Dolha, M., & Beetz, M. (2008). Towards 3D point cloud based object maps for household environments. *Robotics and Autonomous Systems*, *56*(11), 927–941.

Salas Fernandez, M. G., Bao, Y., Tang, L., & Schnable, P. S. (2017). A high-throughput, field-based phenotyping technology for tall biomass crops. *Plant Physiology*, *174*(4), 2008–2022.

Salas Fernandez, M. G., Becraft, P. W., Yin, Y., & Lübberstedt, T. (2009). From dwarves to giants? Plant height manipulation for biomass yield. *Trends in Plant Science*, *14*(8), 454–461.

Sanz, R., Rosell, J. R., Llorens, J., Gil, E., & Planas, S. (2013). Relationship between tree row LIDAR-volume and leaf area density for fruit orchards and vineyards obtained with a LIDAR 3D Dynamic Measurement System. *Agricultural and Forest Meteorology*, *171*, 153–162.

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, *47*(1–3), 7–42.

Shafiekhani, A., Kadam, S., Fritschi, F. B., & DeSouza, G. N. (2017). Vinobot and vinoculer: Two robotic platforms for high-throughput field phenotyping. *Sensors*, *17*(1), 214.

Shi, Y., Thomasson, J. A., Murray, S. C., Pugh, N. A., Rooney, W. L., Shafian, S., … others. (2016). Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PloS One*, *11*(7), e0159781.

Sodhi, P., Vijayarangan, S., & Wettergreen, D. (2017). In-field segmentation and identification of plant structures using 3D imaging. In *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*. Vancouver, Canada: Piscataway, NJ: IEEE.

Truong, S. K., McCormick, R. F., Rooney, W. L., & Mullet, J. E. (2015). Harnessing genetic variation in leaf angle to increase productivity of Sorghum bicolor. *Genetics*, *201*(3), 1229–1238.

Underwood, J. P., Jagbrant, G., Nieto, J. I., & Sukkarieh, S. (2015). Lidar-Based Tree Recognition and Platform Localization in Orchards. *Journal of Field Robotics*, *32*(8), 1056–1074.

Underwood, J. P., Wendel, A., Schofield, B., McMurray, L., & Kimber, R. (2017). Efficient in-field plant phenomics for row-crops with an autonomous ground vehicle. *Journal of Field Robotics*, *34*(6), 1061–1083.

Vijayarangan, S., Sodhi, P., Kini, P., Bourne, J., Du, S., Sun, H., … Wettergreen, D. (2018). High-throughput robotic phenotyping of energy sorghum crops. *Field and Service Robotics*, 99–113.

Virlet, N., Sabermanesh, K., Sadeghi-Tehran, P., & Hawkesford, M. J. (2017). Field Scanalyzer: an automated robotic field phenotyping platform for detailed crop monitoring. *Functional Plant Biology*, *44*(1), 143–153.

Watanabe, K., Guo, W., Arai, K., Takanashi, H., Kajiya-Kanegae, H., Kobayashi, M., … others. (2017). High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. *Frontiers in Plant Science*, *8*, 421.

Yang, Q. (2012). A non-local cost aggregation method for stereo matching. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 1402–1409). Providence, RI: Piscataway, NJ: IEEE.

Yoon, K.-J., & Kweon, I. S. (2006). Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(4), 650–656.

Zabih, R., & Woodfill, J. (1994). Non-parametric local transforms for computing visual correspondence. In *European Conference on Computer Vision* (pp. 151–158). Stockholm, Sweden: Berlin, Heidelberg: Springer.

Zbontar, J., & LeCun, Y. (2016). Stereo matching by training a convolutional neural network to compare image patches. *Journal of Machine Learning Research*, *17*(1–32), 2.

Zhao, J., Mantilla Perez, M. B., Hu, J., & Salas Fernandez, M. G. (2016). Genome-wide association study for nine plant architecture traits in sorghum. *The Plant Genome*, *9*(2).