

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/356402374>

Measuring Stem Diameter of Sorghum Plants in the Field Using a High-Throughput Stereo Vision System

Article in Transactions of the ASABE (American Society of Agricultural and Biological Engineers) · January 2021

DOI: 10.13031/trans.14156

CITATIONS

5

READS

295

4 authors, including:



Lirong Xiang

North Carolina State University

14 PUBLICATIONS 281 CITATIONS

SEE PROFILE



Jingyao Gai

Guangxi University

17 PUBLICATIONS 363 CITATIONS

SEE PROFILE



Le Wang

Zhejiang University

11 PUBLICATIONS 154 CITATIONS

SEE PROFILE

MEASURING STEM DIAMETER OF SORGHUM PLANTS IN THE FIELD USING A HIGH-THROUGHPUT STEREO VISION SYSTEM



Lirong Xiang¹, Lie Tang^{1,*}, Jingyao Gai¹, Le Wang¹

¹ Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, Iowa, USA.

* Correspondence: lietang@iastate.edu.

HIGHLIGHTS

- A custom-built camera module named PhenoStereo was developed for high-throughput field-based plant phenotyping.
- Novel integration of strobe lights facilitated application of PhenoStereo in various environmental conditions.
- Image-derived stem diameters were found to have high correlations with ground truth, which outperformed any previously reported sensing approach.
- PhenoStereo showed promising potential to characterize a broad spectrum of plant phenotypes.

ABSTRACT. *The stem diameter of sorghum plants is an important trait for evaluation of stalk strength and biomass potential, but it is a challenging sensing task to automate in the field due to the complexity of the imaging object and the environment. In recent years, stereo vision has offered a viable three-dimensional (3D) solution due to its high spatial resolution and wide selection of camera modules. However, the performance of in-field stereo imaging for plant phenotyping is adversely affected by textureless regions, occlusion of plants, variable outdoor lighting, and wind conditions. In this study, a portable stereo imaging module named PhenoStereo was developed for high-throughput field-based plant phenotyping. PhenoStereo features a self-contained embedded design, which makes it capable of capturing images at 14 stereoscopic frames per second. In addition, a set of customized strobe lights is integrated to overcome lighting variations and enable the use of high shutter speed to overcome motion blur. PhenoStereo was used to acquire a set of sorghum plant images, and an automated point cloud data processing pipeline was developed to automatically extract the stems and then quantify their diameters via an optimized 3D modeling process. The pipeline employed a **mask region convolutional neural network (Mask R-CNN)** for detecting stalk contours and a semi-global block matching (SGBM) stereo matching algorithm for generating disparity maps. The correlation coefficient (r) between the image-derived stem diameters and the ground truth was 0.97 with a mean absolute error (MAE) of 1.44 mm, which outperformed any previously reported sensing approach. These results demonstrate that, with proper customization, stereo vision can be an effective sensing method for field-based plant phenotyping using high-fidelity 3D models reconstructed from stereoscopic images. Based on the results from sorghum plant stem diameter sensing, this proposed stereo sensing approach can likely be extended to characterize a broad range of plant phenotypes, such as the leaf angle and tassel shape of maize plants and the seed pods and stem nodes of soybean plants.*

Keywords. *Field-based high-throughput phenotyping, Point cloud, Stem diameter, Stereo vision.*

Understanding the relationship between genotype and phenotype plays an important role in plant phenomics (Bao et al., 2019). Plant phenotyping, which refers to the assessment of plant phenotypic features related to growth, tolerance, architecture, and yield, can potentially aid in the identification of high-yielding, stress-tolerant crop species (Li et al., 2014). By dissecting the genetic basis of plant phenotypic traits, robust crop species can be selected by plant breeders. Recent advances in high-throughput genotype screening have offered fast and inexpensive solutions for plant genomics, which has accelerated plant breeding programs (Bao et al., 2019). However, applications of genomic technologies are limited by their ability to accurately and precisely collect of phenotypic traits. Traditional phenotyping procedures are time-consuming, labor-intensive, and low-throughput, and most phenotypes are obtained in a destructive way or involve manual observations (Minervini et al., 2017). Hence, efficient and effective phenotyping strategies are needed to assess plant growth in an automatic and non-invasive way (Furbank and Tester, 2011).

In recent years, various vision-based phenotyping systems have been developed to automate field-based phenotyping.

Previous high-throughput phenotyping systems can be divided into two categories: aerial-based and ground-based measuring

Submitted for review on 13 June 2020 as manuscript number ITSC systems (Li et al., 2014). Aerial-based phenotyping platforms 14156; approved for publication as a Research Article by the Information Technology, Sensors, & Control Systems Community of ASABE on 23 August 2021. can capture large-scale information of plants within minutes. As one of the emerging alternatives in aerial-based platforms,

unmanned aerial vehicles (UAVs) are gaining increased attention due to their low cost, portability, and ease of operation (Barbedo, 2019). UAVs equipped with a variety of sensors have been successfully developed to estimate vegetation indices (Li et al., 2019), canopy temperatures (Perich et al., 2020), biomass accumulation (Devia et al., 2019), and plant height (Lu et al., 2019). However, the use of aerial imaging systems to monitor plant growth has well-known limitations. For example, these systems can only capture top-view images, making it impossible to assess plant traits (e.g., stalk size and leaf angle) that are only observable in side views of plants.

Compared to aerial-based phenotyping platforms, ground-based phenotyping platforms have the advantages of flexible design, large sensor payload, and high sensor resolution. With customized ground vehicles and a wide range of sensors, these systems are capable of traversing a field to measure a rich set of plant traits, such as plant height, leaf area, leaf angle, stalk size, tiller density, and many other organ-level traits (Bao et al., 2018; Baweja et al., 2018; Jin et al., 2019; Sodhi et al., 2017). Some examples of such systems include the Robotanist (Mueller-Sim et al., 2017), the Vinobot (Shafiekhani et al., 2017), and semi-autonomous tractor-based systems (Salas Fernandez et al., 2017). The Robotanist, developed by Carnegie Mellon University, is a field robot that can self-navigate within tightly spaced crop rows and gather phenotypic data with side-facing sensors (Mueller-Sim et al., 2017). The Vinobot is another autonomous field robot with a robotic manipulator to handle multiple sensors at different viewing angles (Shafiekhani et al., 2017).

A variety of imaging sensors, including multispectral, thermal, hyperspectral, and three-dimensional (3D) sensors, have been integrated into phenotyping systems. Among these sensing technologies, 3D sensors have gained great interest for measuring architectural features because of the 3D nature of plant architecture. Light detection and ranging (LiDAR) sensors are widely used 3D sensors for field-based phenotyping. Chakraborty et al. (2019) reconstructed 3D models of apple trees and grapevine canopies by combining 3D LiDAR with an inertial measurement unit (IMU) on a ground vehicle. The canopy volume was derived from the 3D models and was found to be strongly correlated with manual measurements. In addition to canopy-level morphological traits, high-resolution 3D LiDAR sensors have potential for sensing organ-level traits such as cotton bolls (Sun et al., 2020), sorghum panicles (Malambo et al., 2019), and corn stems (Jin et al., 2019). However, such measurements can be complex and require long imaging times to obtain dense and accurate canopy models (Dandrifosse et al., 2020). A time-of-flight (ToF) camera is another widespread sensor that can provide RGB-D data of plants in the field. ToF sensors have been successfully used for plant architecture measurement (Bao et al., 2018), biomass estimation (Krogh et al., 2018), fruit counting (Tu et al., 2020), and weed detection (Li and Tang, 2018). The major weakness of ToF sensors is their sensitiveness to strong sunlight. To alleviate this issue, a shroud or an umbrella could be used to reduce the sunlight intensity in outdoor applications (Li and Tang, 2018).

Stereo vision, which reconstructs a 3D model from multi-view images, offers an inexpensive, accurate, and effective solution for in-field plant phenotyping. Multi-view stereo vision systems combine multiple-view stereo (MVS) and structure-from-motion (SFM) techniques to reconstruct 3D models of plants from images. Nguyen et al. (2016) developed an in-field phenotyping system that used 16 color stereo vision cameras for capturing multi-view images of eggplant. An arc-shaped superstructure was designed to deploy the cameras and structured illumination modules. Plant height, leaf number, leaf area, and plant biomass could be estimated with the system. The MVS-SFM method is capable of generating dense point clouds of plants, but it is challenging to implement in the field, especially for tall-growing plants like corn and sorghum. Binocular stereo vision has been considered as an alternative to overcome the limitations of MVS-SFM. Binocular stereo vision requires just two cameras to compute depth information.

Previous work has demonstrated the potential of stereo vision in measuring plant architectures outdoors. For example, Sodhi et al. (2017) implemented such a system to study sorghum plant architecture in field conditions. In their application, phenotypic traits including leaf length, leaf width, and stem diameter were measured. Bao et al. (2019) developed a stereo-vision-based robotic system for tall dense canopy crops in the field. The stereo cameras were positioned laterally and stacked vertically on an extension rig to perform side-view imaging. Several morphological traits were quantified, and the results were found to be accurate and highly repeatable. However, a wider row spacing was required by the system. Although their system had the advantages of high image resolution, the performance of stereo vision suffered from textureless regions, inadequate spatial resolution, sunlight variations, and wind conditions (Bao et al., 2019; Dandrifosse et al., 2020).

Sorghum is one of the most promising energy crops due to its tolerance of drought and heat stresses (Xin et al., 2008). Stem diameter is an important feature for yield estimation and stalk strength assessment of sorghum plants (Salas Fernandez et al., 2017). Typically, sorghum stem diameter is measured manually using calipers, which is a tedious and error-prone practice, but developing an automated system that can accurately measure the stem diameter of sorghum plants in the field has been challenging (Salas Fernandez et al., 2017). Several studies have investigated the suitability of stereo vision for field-based stem diameter estimation. Bao et al. (2019) developed a stereo vision system to determine stem diameter from side-view images in which the user first clicks on four reference points on the stem edge to identify a representative stem, and the stem diameter is then calculated by the triangulation principle. Although results showed high repeatability and accuracy, the method requires human intervention.

Recent advances in machine learning offer new approaches for reliable object detection in some image-based deep learning tasks. For example, Baweja et al. (2018) used convolutional deep neural networks to detect sorghum stems and calculated the stem width from stereoscopic images. The proposed method was 270 times faster than in-field manual measuring, and a mean absolute error of 19.3% of the average stem width was achieved. These previous studies identified the sorghum stem in 2D images and performed well. However, the performance could be further improved by acquiring higher-quality images.

In addition, 3D geometric features such as surface normal directions, which are robust to varied illumination conditions, can be used to assist in plant organ segmentation (Sun et al., 2020).

Based on the aforementioned literature, this study investigated a robust and accurate machine vision system that can measure the stem diameters of sorghum plants in 3D space in outdoor conditions. The developed vision system was expected to deliver consistent performance in variable lighting and wind conditions in the field and be capable of high-throughput data acquisition for close-range stereoscopic imaging. The specific objectives of this study were to: (1) acquire high-quality images of sorghum plants in the field using a custom-built stereo imaging module, (2) develop an automated image processing pipeline to detect the stems of sorghum plants and measure their diameters in reconstructed 3D models, and (3) evaluate the performance of stem instance segmentation and stem diameter estimation.

MATERIALS AND METHODS

DEVELOPMENT OF STEREO IMAGING SYSTEM

Our custom-built stereo imaging module for plant phenotyping, named PhenoStereo, is capable of capturing high-quality images in the field with strobe lighting. PhenoStereo (fig. 1a) is mainly comprised of a developer kit, two RGB color cameras, a printed circuit board (PCB), and strobe lights. The developer kit includes a Jetson TX2 embedded platform (NVIDIA, Santa Clara, Cal.), an Elroy carrier board (Connect Tech Inc., Guelph, Ont., Canada), and other devices (e.g., SD card and PCIe card) on the carrier board. The Jetson TX2 simultaneously acquires image pairs from the two stereo cameras through high-speed interfaces (PCIe) and saves them to the SD card. The Elroy carrier board was chosen to interface with the Jetson TX2 due to its small size and variety of standard hardware interfaces.

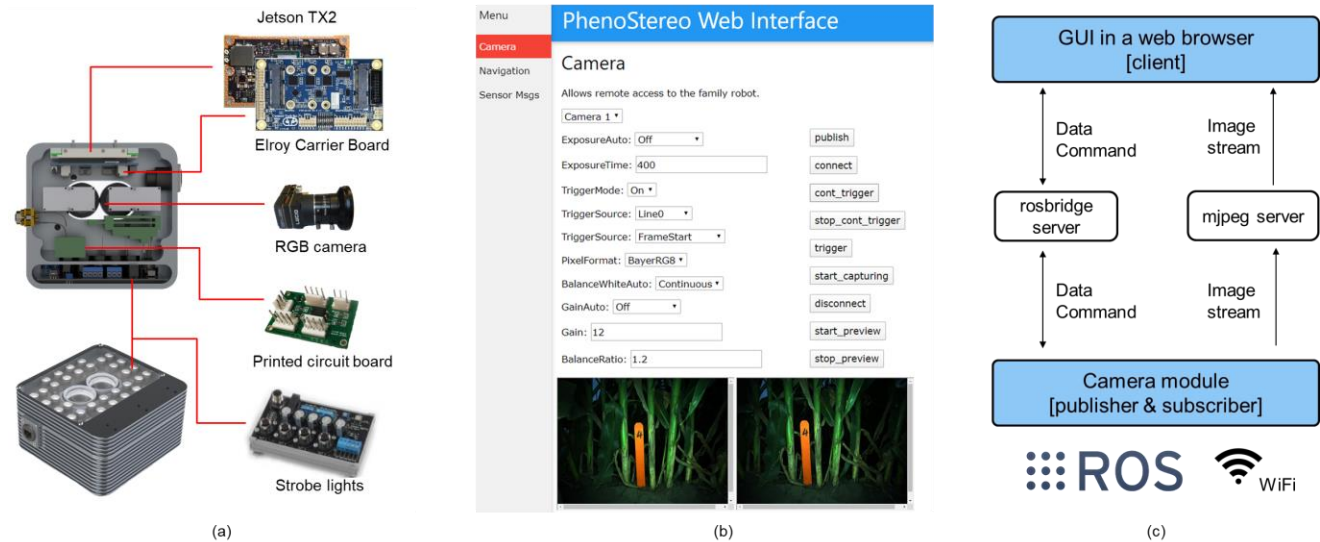


Figure 1. (a) Configuration of PhenoStereo imaging unit, consisting of stereo cameras and strobe lights for capturing high-resolution images, (b) graphic user interface of PhenoStereo, which uses Robot Operating System (ROS) network for wireless control, and (c) software structure.

The two identical RGB cameras (Phoenix 3.2 MP, Lucid Vision Labs, Richmond, B.C., Canada) were each equipped with a lens of 4.0 mm focal length. The baseline of the stereo pair was set at 38 mm, which enabled the module to image close-by objects with a large overlap between the left and right images. The stereo cameras have a horizontal view angle of 85.8° and a vertical view angle of 63.6°. PhenoStereo features onboard storage that performs image acquisition at a maximum rate of 14 frames per second (fps). The raw images, with a resolution of 2048 × 1536 pixels, are stored in the memory of the device and can be easily transferred through an Ethernet interface.

Four sets of high-intensity LEDs (LM75, Smart Vision Lights, Norton Shores, Mich.) were mounted around the camera pair for strobing illumination. The reason for this design was to provide consistent lighting, enhance the visual texture, and allow fast shutter speed. The LED lights were equipped with a 576 W (with 10% duty cycle) driver (CTL-IO-4, Smart Vision Lights) for driving the high-intensity LEDs.

The imaging platform is enclosed in a waterproof metal case (151 mm × 146 mm × 82 mm), and the system weighs approximately 2.06 kg. A transistor switch circuit was designed to synchronize and trigger the camera pair and the strobe lights. The imaging system was able to acquire images with illumination of 31,000 lux at 0.5 m. A web-based graphic user interface (GUI) (fig. 1b) was developed to control the cameras and view the images. The interface allows the user to adjust camera parameters (exposure time, white balance, etc.), send trigger commands, and view live images on a smartphone or laptop browser. The software (fig. 1c) is composed two main modules: (1) the GUI that enables system control and image display, and (2) the camera module that receives and executes the commands. During data collection, the GUI sends the

camera parameters and control commands in Javascript Object Notation (JSON), and the camera module publishes and subscribes ROS messages over a local area network (LAN). The GUI interacts with the camera module wirelessly through Rosbridge (Crick et al., 2017) and MJPEG Server. Rosbridge is composed of two parts: one part is the Rosbridge protocol for sending JSON-based messages and interacting with ROS, and the other part is a collection of packages including a Rosbridge server that provides a WebSocket connection for browsers and the cameras (Coronado and Venture, 2020). To efficiently preview the images, the image stream is transferred in binary format using MJPEG Server. The control algorithm for image acquisition was written in C++, using ROS Kinetic running on Ubuntu 16.04.

DATA COLLECTION

A four-wheel custom-built ground vehicle was used to deploy PhenoStereo (fig. 2). The PhenoStereo system was mounted on the vehicle at a height of 0.51 m (20 in.) above the ground, and aluminum frames were used to support the camera module. The ground vehicle was manually driven at about 3 mph (4.8 km h⁻¹) through 0.76 m (30 in.) spaced crop rows, and images were captured at 10 fps. A smartphone was used to run the web-based GUI to preview the images and adjust the camera parameters. The vehicle was tested on a flat plot with typical cultivated soil conditions. To prevent motion blur caused by vehicle movement and wind conditions, the shutter speed was set to 0.3 ms. The distance between the camera and the sorghum stalks was about 0.38 m (15 in.). Given the camera configuration and the field conditions, the system achieved a vertical field of view (FOV) of approximately 0.47 m and a horizontal FOV of 0.75 m.

Data were acquired at the Agricultural Engineering and Agronomy Research Farm of Iowa State University in Boone, Iowa (fig. 2a). The data collection was performed on September 30 and October 9, 2019. To test PhenoStereo's robustness in varying outdoor lighting conditions, sorghum plant images were collected in sunny (fig. 2b) and overcast (fig. 2c) weather conditions and at different times of day. The camera parameters (white balance, exposure time, and gain) were tuned in the field to produce high-quality images for 3D reconstruction. The ground truth of stem diameter was measured manually using a caliper on stem sections that were 5 to 10 cm above the soil surface on October 9. A total of 75 stems were measured to evaluate the performance of stem diameter estimation. The measured plants were labeled with a tag to make them distinguishable during image processing.

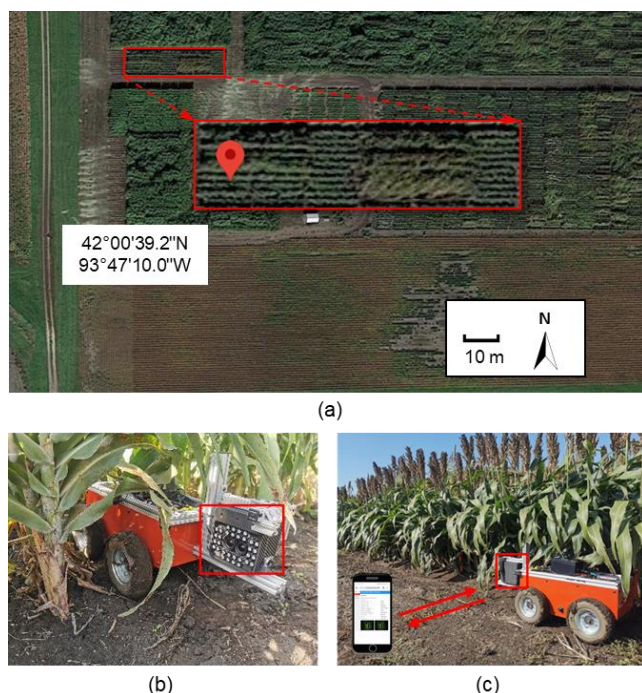


Figure 2. (a) Location of remote-controlled data collection with PhenoStereo in (b) sunny and (c) overcast conditions. A smartphone was used to interact with PhenoStereo using a wireless network.

IMAGE PROCESSING

A series of operations were conducted on the stereo images to extract the stem diameters of sorghum plants, including 3D reconstruction, stem segmentation, point cloud filtering, and stem diameter estimation (fig. 3). First, 3D point clouds of the stems were reconstructed from the stereo images by semi-global block matching (SGBM) (Hirschmüller, 2008). Individual stem masks were then detected using a mask region convolutional neural network (Mask R-CNN) (He et al., 2020), which is a convolutional neural network (CNN) that is capable of instance segmentation at the pixel level. The 2D detections were projected onto a 3D point cloud to segment individual stems. After that, noisy voxels in the segmented point cloud

were detected and filtered using local geometric features. Finally, cylinder fitting was conducted on selected stem sections, and stem diameters were derived from the fitted cylinders. Additional details are provided in the following sections.

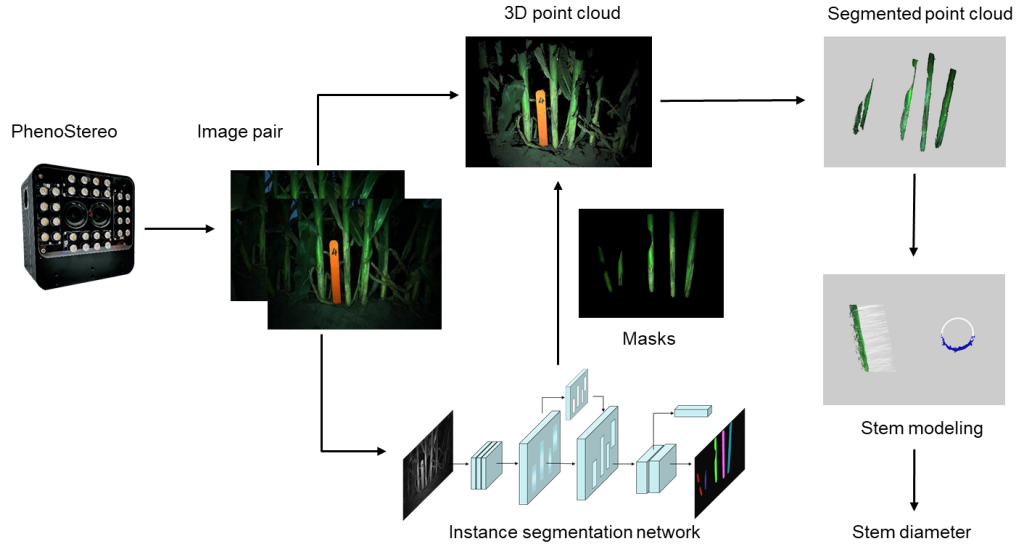


Figure 3. Illustration of image processing pipeline.

3D Reconstruction

The process of 3D reconstruction can be described as follows: first, the parameters of the two cameras were computed by calibration, and the image pair was rectified to enable stereo matching. After that, a stereo matching algorithm was implemented to find the pixel-wise correspondence between the two images to generate a disparity map. The final step was to reconstruct a 3D model from the disparity map by computing the 3D coordinates of the pixels using the triangulation principle and camera parameters.

The camera calibration was performed with a checkerboard pattern (Bradski and Kaehler, 2008) to compute the intrinsic and extrinsic parameters of the stereo cameras. With the camera intrinsic parameters, the image pair was rectified by applying a projective transformation. Because the images followed the principle of epipolar constraints (Brown et al., 2003), two corresponding points are on the same horizontal line in the rectified image pair (fig. 4a). The rectification process removed the lens distortion and enabled reduction of the search space of the stereo correspondence from 2D to 1D.

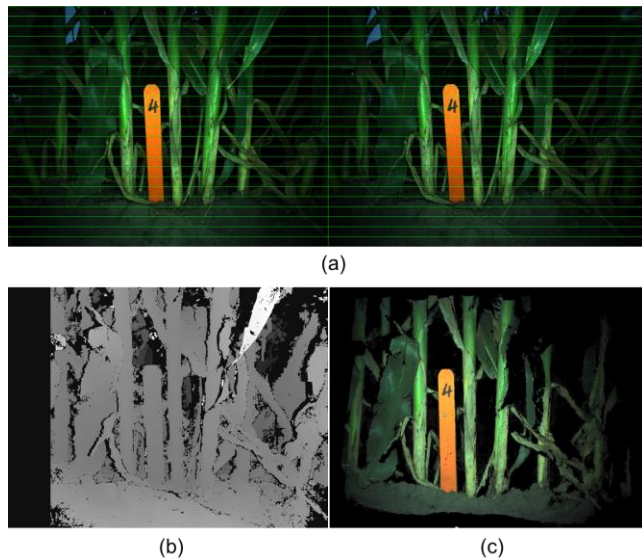


Figure 4. Stereo matching using semi-global block matching (SGBM): (a) rectified and segmented image pair in which corresponding points are on the same horizontal scan lines, (b) disparity output of SGBM algorithm, and (c) 3D point cloud reconstruction of sorghum plant.

To extract the 3D coordinates of the object, the rectified image pair was used to generate a disparity map by stereo matching. In stereo matching of field crop images, two challenges can arise. The first challenge is the nearly homogeneous

color and texture of plant canopy surfaces, which can make the correspondence search ambiguous. The second challenge is occlusion, which means that some pixels in one image do not have correspondences in the other image. In this study, SGBM (Hirschmüller, 2008) was adopted for stereo matching. The principle of SGBM is to perform scan line optimization along multiple directions and aggregate the matching cost from each direction to enforce the smoothness constraint. Although this method is not as accurate as other advanced algorithms, it produces competitive results with better computational efficiency (Bao et al., 2019). SGBM has also shown great potential in retaining edges and dealing with untextured areas (Hirschmüller, 2008), which is important for 3D reconstruction of in-field crop plants. Considering that PhenoStereo has a short baseline (38 mm) and the sorghum plants were about 38 cm (15 in.) away from the camera, the matching window size and the disparity range were set to 5 and 100 pixels, respectively. The result of stereo matching is the disparity map (fig. 4b), which gives the apparent pixel difference in the left and right images. The disparities are inversely proportional to the depths.

After working out the disparities of the images, the 3D coordinates (X , Y , Z) can be extracted from the images and the disparity map based on a triangulation method. We considered the rectified left image as a reference, and the X - Y coordinates and the distance between the object and the camera (Z) are given by equation 1:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{x \times b}{d} \\ \frac{y \times b}{d} \\ \frac{f \times b}{d} \end{bmatrix} \quad (1)$$

where x and y are the coordinates of a pixel in the 2D image, d is the corresponding disparity value, b is the baseline (m) of the stereo camera, and f is the focal length (pixel). The focal length and baseline can be obtained in the calibration process. After stereo matching, a 3D point cloud can be obtained (fig. 4c). Points with z coordinates greater than 0.8 m or less than 0.002 m were removed as background or noise. The obtained 3D model was processed to determine the stem diameter according to the geometric features.

Stem Segmentation

For individual stem measurement, effective stem segmentation is the prerequisite for further image analysis. Identifying the stems in the plant canopy can be difficult because of the color and texture similarities between plant organs. The occlusions created by the dense canopy of sorghum plants also pose challenges for automated segmentation of stems. In this study, we first detected the individual stems in 2D images. The 2D detections were then rectified and projected onto the 3D point cloud to segment the stems.

Mask R-CNN (He et al., 2020) was used to detect and segment sorghum stems from the RGB images (fig. 5). The model first employs a feature pyramid network (FPN) (Lin et al., 2017) to construct feature maps of objects, and then feeds the feature maps into the region proposal network (RPN) to generate regions of interest (ROIs). Target features are extracted and mapped for ROIs by the RoiAlign layers and sent to a fully convolutional network (FCN) for classification prediction and instance segmentation. According to the elongated shape of the target (sorghum stems), three length-to-width ratios (2:1, 4:1, and 8:1) of the ROI proposals were used to improve the computational efficiency.

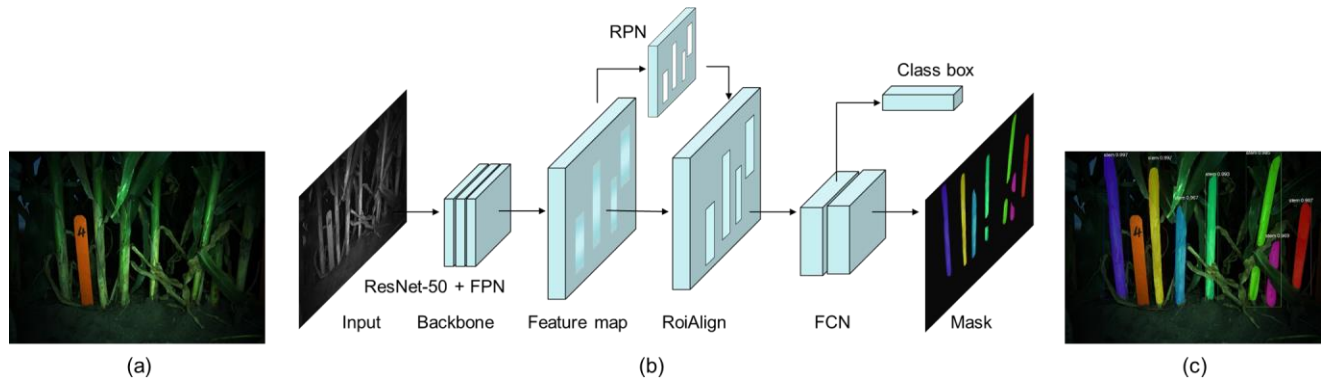


Figure 5. Illustration of Mask R-CNN based stem segmentation: (a) example of input RGB image, (b) diagram of Mask R-CNN architecture, and (c) masked image with detected stems.

The detection of individual stems was formulated as a semantic segmentation task with two classes: stem and background. Pixels of the stem class referred to the stalks of sorghum plants, while the remaining pixels were assigned to the background class. We manually annotated 440 images, in which each plant stem was colored independently. The labeled images were split into a training set, a testing set, and a validation set with a proportion of 7:2:1. Although the strobe lights were used to maintain consistent lighting, the images varied in color rendering because of different camera settings, such as auto white

balance mode. Therefore, to alleviate the variations in color between images, the RGB images were converted to gray-scale images. This process reduced the image dimensions from three channels to one channel, which also decreased the complexity of the network architecture.

The Mask R-CNN algorithm was implemented on the TensorFlow framework. In this study, the ResNet network was used as the backbone for the feature extractor due to its great potential for alleviating gradient disappearance and training degradation (Yu et al., 2019). Two ResNet backbone networks were tested for the model, including ResNet-50 and ResNet-101. The model trained with ResNet-50 achieved better performance on the testing dataset; therefore, ResNet-50 was used as the backbone network for stem segmentation. The Mask R-CNN model was initialized using pretrained weights from the COCO dataset (Lin et al., 2014). Image augmentation was used to increase the number of training images and diminish overfitting. The image augmentation allowed the networks to be more adaptive to field conditions. We applied crop, flip, Gaussian blur, multiply, and affine transformation to simulate the differences in real-world cases, such as the variations caused by camera settings, lighting conditions, object-to-camera distance, and noise (Jung et al., 2019 [Not listed in the References]). During the training process, the augmentations were randomly assigned to 50% of the source images to enlarge the training dataset artificially. Based on the preliminary experiments, we trained the Mask R-CNN model using an initial learning rate of 0.001, a learning momentum of 0.9, and a weight decay of 0.0001. A stochastic gradient descent (SGD) algorithm was implemented for the network training. The model was trained for a total of 200 epochs with a batch size of 8. The obtained masks were applied to the original RGB image for segmentation of individual stems (fig. 5c).

Point Cloud Filtering

Filtering was one of the most important steps in the proposed algorithm because the accuracy of the cylinder fitting and stem diameter estimation depended on it. The filtering process aimed to eliminate everything that did not belong to the cylindrical stem, including the random noise introduced by stereo matching and the non-stem points (such as leaf collars) that were misclassified by the instance segmentation process.

During the filtering process, we denoted the stem points and non-stem points as inliers and outliers, respectively. We noticed that most of the outliers were from leaf collars (fig. 6a), which is the intersection between a leaf and a stem. For the filtering process, color and/or intensity values were unreliable because some outliers had almost the same color as the inliers. Geometrically, the inliers and outliers exhibited distinguishing local features, which can be attributed to (1) the cylindrical structure of the stems, and (2) the relatively smooth and continuous surface of the stems. Based on that observation, we opted to use local surface features to highlight the stem points. The point cloud was stored in a k - d tree, which is a binary search tree that enables fast lookup of range and nearest neighbors. Principal component analysis (PCA) was implemented to characterize the shape features of each individual point. For a point p in the point cloud, the k -neighborhood point set can be defined as \mathcal{N}_k (eq. 2), where d is the search radius of neighbor points in 3D space. The covariance matrix (C) of region \mathcal{N}_k can be computed using equation 3, where \bar{p} (eq. 4) denotes the centroid of the region. We then performed PCA on C to obtain the eigenvalues λ_1 , λ_2 , and λ_3 ($\lambda_1 > \lambda_2 > \lambda_3$):

$$\mathcal{N}_k = \{p_i: p - p_i < d\} \quad (2)$$

$$C = \frac{1}{k} \sum_{i=1}^k (p_i - \bar{p}) \cdot (p_i - \bar{p})^T \quad (3)$$

$$\bar{p} = \frac{1}{k} \sum_{i=1}^k p_i \quad (4)$$

Different arithmetic combinations of the eigenvalues have been proposed as 3D local feature descriptors for point cloud classification (Hackel et al., 2016). For example, if the three eigenvalues are close to each other ($\lambda_1 \approx \lambda_2 \approx \lambda_3$), the point is part of a region that has an isotropic spatial distribution, such as a spherical structure. For a cylindrical structure such as a sorghum stem, we expect λ_3 to be much smaller than λ_1 and λ_2 ($\lambda_1 > \lambda_2 \gg \lambda_3$). In this study, cylindricity was defined as a structure tensor to discriminate stem and non-stem points. The cylindricity (c) at a point p is expressed by equation 5:

$$c = \frac{\lambda_1 - \lambda_2}{\lambda_1 - \lambda_3} \quad (5)$$

Numerically, a point and its neighbor points belonging to a sorghum stem should have small c values. The cylindricity value of each point was chosen to eliminate leaf collars and noise points from the stem. We binarized the point cloud into stem and non-stem points based on Otsu's method (Otsu, 1979) (fig. 6b), and the points with c values greater than the Otsu threshold were removed as outliers (fig. 6c).

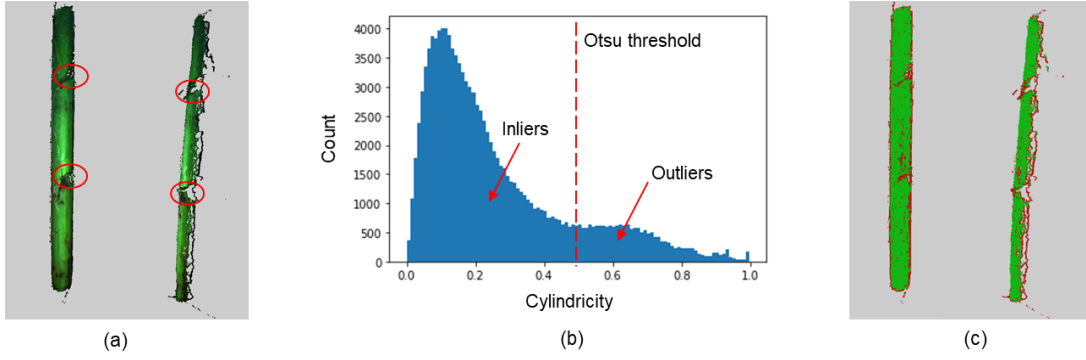


Figure 6. Point cloud filtering using local features: (a) point cloud segment of a stem (leaf collars are highlighted with red circles), (b) histogram of cylindrical difference in stem and non-stem points (red dotted line is the Otsu threshold, which separates stem inliers from noise), and (c) result from proposed filtering algorithm, where green points are stem inliers and red points are outliers.

Stem Diameter Estimation

Fitting a cylindrical shape for the plant stem proved to be a reliable modeling method (Chaivivatrakul et al., 2014). Therefore, we modeled sorghum stems as cylindrical shapes for stem diameter estimation. However, cylinder fitting can be a computationally intensive process because of its high parametric space. To reduce the complexity and dimension of the parameter space, we split the process into first finding the cylindrical axis direction and then detecting a circle in a 2D plane. The radius of the circle was defined as the stem diameter in this study.

Surface normals are important properties for analyzing a geometric surface. The normal of a point P is a vector that is perpendicular to the tangent plane at P (fig. 7a). It is not hard to determine that the normal vectors of a cylinder are perpendicular to its axis. Mathematically, the cross product of the surface normal and the cylinder axis direction vector is supposed to be zero. Based on this principle, normal estimation was implemented to estimate the orientation of the cylindrical axis. In this study, the normal vector of a given point was computed by finding the eigenvector corresponding to λ_3 , which is the smallest eigenvalue of the covariance matrix in equation 3. We calculated the surface normal for each point in the point cloud (figs. 7b and 7c) to find a vector that was most orthogonal to the surface normals of all the points. The orientation of the cylinder axis was defined as $\vec{\omega}(a,b,1)$, and the parameters were determined by least-squares method using the cost function $\mathcal{F}(a,b)$ given by equation 6:

$$\begin{aligned}\mathcal{F}(a,b) &= \sum_{i=1}^m (\vec{\omega} \cdot \vec{n}_i)^2 \\ &= \sum_{i=1}^m (a \cdot x_i + b \cdot y_i + z_i)^2\end{aligned}\quad (6)$$

where $\vec{n}_i(x_i, y_i, z_i)$ is the normal vector of the i th point in the point cloud, and m is the number of points.

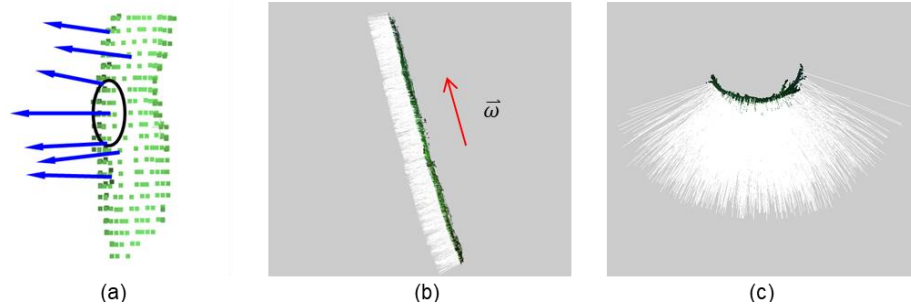


Figure 7. Estimation of direction of cylindrical axis: (a) illustration of surface normal estimation, (b) side view of surface normals of a stem (vector $\vec{\omega}$ is the direction of the cylindrical axis, and white epidermal hairs are surface normals), and (c) top view of surface normals of the stem.

Before projecting the point cloud onto a 2D plane, the stem was rotated so that the cylindrical axis was aligned with the z coordinates. The direction vector of the cylinder axis ($\vec{\omega}$) was normalized as $\vec{\omega}_0$. The unit vector $\vec{u}_z(0,0,1)$ was the z -axis vector. At first, we computed the cross product (eq. 7) and the dot product (eq. 8) of the two vectors. According to Rodrigues' rotation formula, the rotation matrix (eq. 9) from vector $\vec{\omega}_0$ to vector $\vec{u}_z(0,0,1)$ can be obtained:

$$\vec{v} = \vec{\omega}_0 \times \vec{u}_z \quad (7)$$

$$k = \vec{\omega}_0 \cdot \vec{u}_z \quad (8)$$

$$R = I + [\sim \bar{v}] + [\sim \bar{v}]^2 \cdot \frac{1}{1+k} \quad (9)$$

where I is the identity matrix, and $[\sim \bar{v}]$ is the skew-symmetric matrix of \bar{v} :

$$[\sim \bar{v}] = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix} \quad (10)$$

Finally, the stem point cloud was rotated by matrix R to ensure that the cylindrical axis was parallel to the z -axis (fig. 8a). After rotation, the points with z values from -15 to -10 cm, which approximately represented the stem section at 5 to 10 cm above the ground, were selected for stem diameter estimation. The selected point cloud was projected onto the x - y plane, where the projected points were distributed as a partial circle (fig. 8b). The center (x_0, y_0) and radius (R) of the circle were optimized by performing an unconstrained minimization of $\mathcal{F}(x_0, y_0, R)$ in the three-dimensional parameter space (eq. 11). The Levenberg-Marquardt algorithm (LMA) (Levenberg, 1944; Marquardt, 1963) was used for the optimization process:

$$\begin{aligned} \mathcal{F}(x_0, y_0, R) &= \sum_{i=1}^n d_i^2 \\ &= \sum_{i=1}^n \left(\sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} - R \right)^2 \end{aligned} \quad (11)$$

where d_i is the Euclidean distance from a point to the circle center, and (x_i, y_i) are the x - y coordinates of the i th point in the point cloud.

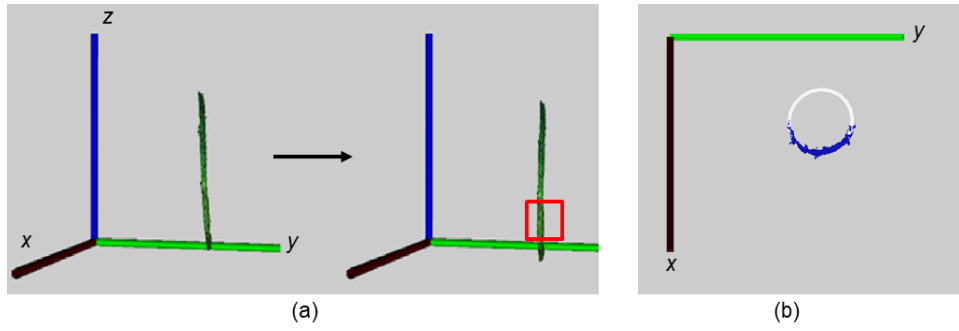


Figure 8. Circle detection using Levenberg-Marquardt algorithm (LMA): (a) 3D transformation of the stem point cloud (the point cloud was rotated to align its cylindrical axis parallel with the z -axis, and the stem section in the red square was selected for circle fitting), and (b) circle fitting in the x - y plane (blue points are the projected point cloud of the selected stem section, and the white circle is the LMA output).

ACCURACY ASSESSMENT

To evaluate the performance of stem segmentation, three widely used evaluation metrics were chosen, including precision (P), recall (R), and harmonic mean (F_1), which can be defined as:

$$P = \frac{TP}{TP + FP} \quad (13)$$

$$R = \frac{TP}{TP + FN} \quad (14)$$

$$F_1 = \frac{2 \cdot P \cdot R}{P + R} \quad (15)$$

where TP is the number of correctly segmented stem pixels (true positives), FP is the number of background pixels incorrectly classified as stem pixels (false positives), and FN is the number of non-segmented stem pixels (false negatives). In addition, the intersection over union (IoU) metric was used as an indicator of position accuracy. IoU is the intersection area of the predicted mask and the ground truth divided by their union area.

The performance of stem diameter estimation was evaluated by performing linear regression analyses between the system-derived measurements and the ground truth. We computed the following statistics to assess the model: Pearson correlation coefficient (r), root mean square error (RMSE, eq. 16), and mean absolute error (MAE, eq. 17).

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_i^N (d_i - d_i^{gt})^2} \quad (16)$$

$$\text{MAE} = \frac{1}{N} \sum_i^N |d_i - d_i^{gt}| \quad (17)$$

where N is the total number of stems used for evaluation, d_i is the image-derived stem diameter of the i th stem, and d_i^{gt} is its corresponding ground truth.

RESULTS

PERFORMANCE OF STEM SEGMENTATION

The testing dataset was used to further evaluate the trained Mask R-CNN model and weights for instance segmentation, where the confidence threshold was set as 0.9 for the stem class. We determined the precision, recall, and F_1 values over a range of IoU thresholds from 0.4 to 0.7 (table 1). For IoU values less than or equal to 0.5, all the precision, recall, and F_1 values were greater than 0.8, which met the requirement for stem instance segmentation in this study. Furthermore, the precision, which represents the ratio of correctly classified pixels in all returned pixels, was impressive even with an IoU value of 0.7. For each IoU value, it is obvious that the recall rate was lower than the precision rate. This was mainly because of non-detected stem pixels close to the soil, where the texture and color were not prominent enough. However, because we only measured the diameters of stem sections at 5 to 10 cm above the ground, the misdetection of pixels near the soil did not affect the stem diameter estimation performance. Overall, the proposed Mask R-CNN model produced satisfactory results for instance segmentation (fig. 9). Despite the relatively small size of the training images, the model was adaptive to variable texture, stem size, occlusion, and illumination conditions.

Table 1. Results of instance segmentation by Mask R-CNN.

IoU	Precision	Recall	F_1
0.4	0.97	0.82	0.87
0.5	0.96	0.81	0.86
0.6	0.93	0.76	0.82
0.7	0.84	0.65	0.71



Figure 9. Example segmentation results produced by Mask R-CNN model. Original images are shown in the top row, and the corresponding segmentation results are shown in the bottom row. Each color area indicates a segmented stem.

PERFORMANCE OF STEM DIAMETER ESTIMATION

The image-derived stem diameters were found to be highly correlated ($r = 0.97$) and accurate ($\text{RMSE} = 1.39$ mm) with respect to the manually measured stem diameters (fig. 10). The proposed method was also highly repeatable, with a standard deviation of 1.76 mm. Considering the range of the measured stem diameters, the achieved MAE (1.44 mm) equaled 5.67% of the ground truth stem diameters (minimum = 9.3 mm, maximum = 34.67 mm). Errors in system-derived diameters were attributable to several factors, including falsely detected stem edges, inaccurate stereo matching, and imperfect camera calibration. The calibration errors resulted in inaccurate estimation of the extrinsic as well as intrinsic camera parameters. The fitted line was close to the reference line ($y = x$), with an intercept of 0.53 mm. The intercept demonstrated that the image-

derived diameter was slightly larger than the corresponding ground truth. This was mainly caused by the soft leaf sheath around the stems, as the caliper tended to give a smaller value when tightened across the stems.

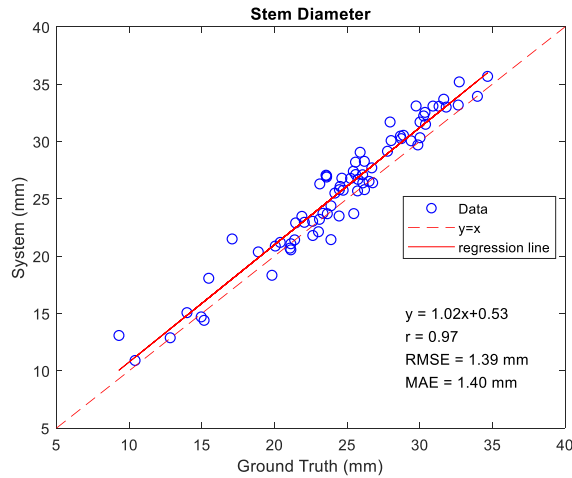


Figure 10. Correlation of stem diameters estimated using Mask R-CNN system and manually measured ground truth.

The proposed method showed lower reliability when the stem diameter was less than 20 mm. A possible explanation is the inherent difficulty in reconstructing thin structures in the process of stereo matching. Furthermore, the cross-section of the stem is best modeled as an ellipse, especially when the stem is small; therefore, the process of stem modeling and the placement of the caliper can introduce errors.

Overall, the method proposed in this study outperformed current automated and user-interactive methods. The automated StalkNet method (Baweja et al., 2018) showed promising efficiency in the measurement of stem width, but with an MAE (2.77 mm) equal to 19.3% of the average stem width. The user-interactive method developed by Bao et al. (2019) showed comparable performance ($r = 0.95$, $MAE = 1.64 \text{ mm}$) with our automated method, but the intercept (2.84 mm) was substantially larger than that of our method. Additionally, the method developed by Bao et al. (2019) requires human intervention to assist in the segmentation of stems, which is laborious and impractical for large-scale studies. Most of the reported methods computed the metric width of a stem from its pixel width in a 2D image. To the best of our knowledge, this is the first study using high-accuracy 3D surface models to estimate the stem diameters of field-grown sorghum plants.

COMPUTATIONAL EFFICIENCY

The image processing pipeline was run on a desktop workstation with a 2.2 GHz Xeon Gold 5120 CPU, 32 GB RAM, and an NVIDIA Titan Xp GPU. The instance segmentation network was trained on an Nvidia GTX Titan X GPU with a 3.5 GHz Xeon HexaCore CPU and 16 GB RAM. Table 2 summarizes the computation time for each processing stage. The image processing consisted of two parts (data extraction and network training), and data extraction comprised four processing stages: stereo matching, stem detection, point cloud filtering, and stem diameter estimation. Usually, the time required for point cloud processing is closely linked to the size of the point cloud. The time consumed by the proposed algorithm mainly occurred during point cloud filtering and was 49% of the total processing time. During the filtering process of a dense point cloud, traversing each point to search for its neighbors incurs much computation time. However, the stem diameter estimation, operating on the point cloud, was fast, and the average computation time was 3.8 s per stem. After evaluating ten representative stems, the average time needed for data extraction of a stem was about 20.9 s, which is much faster and less laborious than in-field manual measurement.

Table 2. Average computation time for each processing stage.

Processing Stage	Average Time
Stereo matching	5.4 s per image
Stem detection	1.4 s per image
Point cloud filtering	10.3 s per stem
Stem diameter estimation	3.8 s per stem
Mask R-CNN training	35 h

DISCUSSION

PERFORMANCE OF PHENOSTEREO

Development of an imaging system that is robust in field conditions is challenging. The major challenges of field-based imaging include variable illumination, wind and weather conditions, and complex backgrounds. In this study, the customized

strobe lighting overcame the challenges of various lighting conditions, and the high-powered flash combined with faster shutter speed reduced the adverse effects of motion due to wind. Furthermore, with lighting of the viewed object, PhenoStereo was capable of producing images with strong contrast between the foreground and background.

Field-based phenotyping is an important alternative to UAV-based phenotyping due to its ability to determine organ-level phenotypic traits at the middle and bottom of the plant canopy. However, most current stereo-vision-based phenotyping platforms have been reported to work best in uniform lighting and in the absence of wind (Dandrifosse et al., 2020; Kaczmarek, 2017; Wu et al., 2020). With its customized strobe lighting system, PhenoStereo shows great potential for overcoming these limitations. Moreover, PhenoStereo is equipped with a ROS-based API and a Gigabit Ethernet interface, making it highly integrative with many field-based phenotyping systems. Compared with manual imaging, the speed and efficiency of data acquisition make PhenoStereo feasible for large-scale phenotyping.

POTENTIAL IMPROVEMENTS IN THE FUTURE

There is still room for improving both the stereo vision system and the image processing pipeline. To improve the stereo vision system, polarizing filters can be added to the strobe lights and camera lenses to reduce noise caused by glare. The stem and leaf surfaces of plants tend to be reflective, especially when the camera is too close to the viewed object, resulting in a negative effect on image quality (fig. 11). Another possible improvement is to use more than one camera set to build a stereo vision system. With multiple adjacent cameras, the system could have a wider field of view, allowing some cameras to have advantageous view angles for areas containing partial objects. A multi-camera system has potential to generate more accurate distance estimations and better view angles over occluded areas (Kaczmarek, 2017).



Figure 11. Partially over-exposed images due to reflective surfaces of sorghum stems.

Concerning the image processing pipeline, more advanced stereo matching algorithms could be used to improve the quality of the disparity maps. This study focused on SGBM for computing stereo correspondence due to its high efficiency and sufficient performance. However, SGBM also has limitations. In this study, the surface reconstructed from SGBM was found to be blurry near object boundaries, which gave inaccurate contours of stems and leaves. This is because the SGBM's filter-based matching cost aggregation tended to smooth the depth discontinuities, and hence cause distortions at image edges (Hirschmüller and Scharstein, 2007). Many advanced techniques have been proposed to improve the edge preservation, including optimization-based filters and weighted-average based smoothing approaches (Zhang et al., 2015). On the other hand, state-of-the-art CNNs could be used to further improve the process of computing the similarity between stereo image patches. For instance, Bao et al. (2019) implemented a robust stereo matching algorithm, 3DMST (Li et al., 2017), to reconstruct surface models of dense plant canopies in the field. The 3DMST algorithm uses CNNs for computing matching costs and nonlocal support region filtering for preserving edges. Although with lower computational efficiency, 3DMST outperformed SGBM in surface smoothness and edge preservation for the surface reconstruction of sorghum plants.

In this study, disparity estimation was performed on both the foreground and background, although the background is not needed for stem diameter estimation. The computational efficiency could be improved by performing disparity estimation only for detected stem regions. In the image processing pipeline, the most time-consuming processing stage was point cloud filtering, mostly due to the expensive PCA analysis for each point. A possible solution to improve the efficiency could be implementing a down-sampling process to reduce the size of the point cloud before filtering.

CONCLUSIONS

This study proposed a stereo vision system, named PhenoStereo, for field-based high-throughput plant phenotyping. The system is capable of producing high-quality and high-sharpness stereoscopic images with fast shutter speed. The novel integration of strobe lights facilitates application of PhenoStereo in various environmental conditions (direct sunlight, back-lighting conditions, shadows, wind conditions). As a case study, we used PhenoStereo to quantify the stem diameters of sorghum plants. The high-quality stereo images allowed construction of accurate surface models of plant stems. Subsequently, an automated image processing pipeline was developed to segment individual stems and perform modeling on the segmented point cloud. The correlation coefficient (r) between the image-derived and ground truth measurements of stem diameter was 0.97 with an MAE of 1.44 mm, surpassing the best values reported in the literature.

To conclude, the proposed method offers an automated, rapid, and accurate solution for extracting the stem diameters of sorghum plants. Our study also demonstrated that, with proper customization, stereo vision is feasible for 3D-based plant phenotyping in field conditions. Future work will focus on improving the quality of the disparity maps, as well as using PhenoStereo to characterize other organ-level morphological traits, such as leaf angle and panicle or tassel size.

ACKNOWLEDGEMENTS

This project was funded by the National Science Foundation Major Research Instrumentation Program (Grant No. DBI-1625364) and the Plant Sciences Institute at Iowa State University. We thank Taylor Tuel, Benjamin Steward, and Duan Liu for their assistance in hardware design; and Juan Pabelo for his contributions to the collection of ground truth data.

REFERENCES

- Bao, Y., Tang, L., Breitzman, M. W., Salas Fernandez, M. G., & Schnable, P. S. (2019). Field-based robotic phenotyping of sorghum plant architecture using stereo vision. *J. Field Rob.*, 36(2), 397-415. <https://doi.org/10.1002/rob.21830>
- Bao, Y., Tang, L., Srinivasan, S., & Schnable, P. S. (2018). Field-based architectural traits characterisation of maize plant using time-of-flight 3D imaging. *Biosyst. Eng.*, 178, 86-101. <https://doi.org/10.1016/j.biosystemseng.2018.11.005>
- Barbedo, J. G. (2019). A review on the use of unmanned aerial vehicles and imaging sensors for monitoring and assessing plant stresses. *Drones*, 3(2), article 40. <https://doi.org/10.3390/drones3020040>
- 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. In M. Hutter & R. Siegwart (Eds.), *Field and service robotics* (pp. 271-284). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-67361-5_18
- Bradski, G. R., & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library* (1st Ed.). Sebastopol, CA: O'Reilly Media.
- Brown, M. Z., Burschka, D., & Hager, G. D. (2003). Advances in computational stereo. *IEEE Trans. Pattern Anal. Mach. Intel.*, 25(8), 993-1008. <https://doi.org/10.1109/TPAMI.2003.1217603>
- Chaivivatrakul, S., Tang, L., Dailey, M. N., & Nakarmi, A. D. (2014). Automatic morphological trait characterization for corn plants via 3D holographic reconstruction. *Comput. Electron. Agric.*, 109, 109-123. <https://doi.org/10.1016/j.compag.2014.09.005>
- Chakraborty, M., Khot, L. R., Sankaran, S., & Jacoby, P. W. (2019). Evaluation of mobile 3D light detection and ranging based canopy mapping system for tree fruit crops. *Comput. Electron. Agric.*, 158, 284-293. <https://doi.org/10.1016/j.compag.2019.02.012>
- Coronado, E., & Venture, G. (2020). Towards IoT-aided human-robot interaction using NEP and ROS: A platform-independent, accessible and distributed approach. *Sensors*, 20(5), article 1500. <https://doi.org/10.3390/s20051500>
- Crick, C., Jay, G., Osentoski, S., Pitzer, B., & Jenkins, O. C. (2017). Rosbridge: Ros for non-Ros users. In *Robotics research* (pp. 493-504). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-29363-9_28
- Dandrifosse, S., Bouvry, A., Leemans, V., Dumont, B., & Mercatoris, B. (2020). Imaging wheat canopy through stereo vision: Overcoming the challenges of the laboratory to field transition for morphological features extraction. *Front. Plant Sci.*, 11, article 96. <https://doi.org/10.3389/fpls.2020.00096>
- Devia, C. A., Rojas, J. P., Petro, E., Martinez, C., Mondragon, I. F., Patino, D., ... Colorado, J. (2019). High-throughput biomass estimation in rice crops using UAV multispectral imagery. *J. Intel. Robotic Syst.: Theory Appl.*, 96(3-4), 573-589. <https://doi.org/10.1007/s10846-019-01001-5>
- Furbank, R. T., & Tester, M. (2011). Phenomics-technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.*, 16(12), 635-644. <https://doi.org/10.1016/j.tplants.2011.09.005>
- Hackel, T., Wegner, J. D., & Schindler, K. (2016). Contour detection in unstructured 3D point clouds. *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (pp. 1610-1618). <https://doi.org/10.1109/CVPR.2016.178>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2020). Mask R-CNN. *IEEE Trans. Pattern Anal. Machine Intel.*, 42(2), 386-397. <https://doi.org/10.1109/TPAMI.2018.2844175>
- Hirschmüller, H. (2008). Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Machine Intel.*, 30(2), 328-341. <https://doi.org/10.1109/TPAMI.2007.1166>
- Hirschmüller, H., & Scharstein, D. (2007). Evaluation of cost functions for stereo matching. *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (pp. 1-8). <https://doi.org/10.1109/CVPR.2007.383248>
- Jin, S., Su, Y., Wu, F., Pang, S., Gao, S., Hu, T., ... Guo, Q. (2019). Stem-leaf segmentation and phenotypic trait extraction of individual maize using terrestrial LiDAR Data. *IEEE Trans. Geosci. Remote Sensing*, 57(3), 1336-1346. <https://doi.org/10.1109/TGRS.2018.2866056>
- [Add "Jung et al., 2019" as cited on page 7]
- Kaczmarek, A. L. (2017). Stereo vision with equal baseline multiple camera set (EBMCS) for obtaining depth maps of plants. *Comput. Electron. Agric.*, 135, 23-37. <https://doi.org/10.1016/j.compag.2016.11.022>
- Krogh, A., Bender, A., Whelan, B., Barbour, M. M., Sukkari, S., Karstoft, H., & Gislum, R. (2018). Segmentation of lettuce in coloured 3D point clouds for fresh weight estimation. *Comput. Electron. Agric.*, 154, 373-381. <https://doi.org/10.1016/j.compag.2018.09.010>
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Q. J. Appl. Math.*, 2(2), 164-168. <https://doi.org/10.1090/qam/10666>
- Li, J., & Tang, L. (2018). Crop recognition under weedy conditions based on 3D imaging for robotic weed control. *J. Field Rob.*, 35(4), 596-611. <https://doi.org/10.1002/rob.21763>
- Li, J., Veeranampalayam-Sivakumar, A.-N., Bhatta, M., Garst, N. D., Stoll, H., Stephen Baenziger, P., ... Shi, Y. (2019). Principal variable selection to explain grain yield variation in winter wheat from features extracted from UAV imagery. *Plant Methods*, 15(1), article 123. <https://doi.org/10.1186/s13007-019-0508-7>

- Li, L., Yu, X., Zhang, S., Zhao, X., & Zhang, L. (2017). 3D cost aggregation with multiple minimum spanning trees for stereo matching. *Appl. Opt.*, 56(12), 3411-3420. <https://doi.org/10.1364/AO.56.003411>
- Li, L., Zhang, Q., & Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors*, 14(11), 20078-20111. <https://doi.org/10.3390/s141120078>
- Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (pp. 936-944). <https://doi.org/10.1109/CVPR.2017.106>
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Proc. European Conf. Computer Vision* (pp. 740-755). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-10602-1_48
- Lu, N., Zhou, J., Han, Z., Li, D., Cao, Q., Yao, X., ... Cheng, T. (2019). Improved estimation of aboveground biomass in wheat from RGB imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system. *Plant Methods*, 15(1), article 17. <https://doi.org/10.1186/s13007-019-0402-3>
- Malambo, L., Popescu, S. C., Horne, D. W., Pugh, N. A., & Rooney, W. L. (2019). Automated detection and measurement of individual sorghum panicles using density-based clustering of terrestrial lidar data. *ISPRS J. Photogram. Remote Sens.*, 149, 1-13. <https://doi.org/10.1016/j.isprsjprs.2018.12.015>
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.*, 11(2), 431-441. <https://doi.org/10.1137/0111030>
- Minervini, M., Giuffrida, M. V., Perata, P., & Tsaftaris, S. A. (2017). Phenotiki: an open software and hardware platform for affordable and easy image-based phenotyping of rosette-shaped plants. *Plant J.*, 90(1), 204-216. <https://doi.org/10.1111/tpj.13472>
- Mueller-Sim, T., Jenkins, M., Abel, J., & Kantor, G. (2017). The Robotanist: A ground-based agricultural robot for high-throughput crop phenotyping. *Proc. IEEE Intl. Conf. Robotics and Automation* (pp. 3634-3639). <https://doi.org/10.1109/ICRA.2017.7989418>
- Nguyen, T. T., Slaughter, D. C., Townsley, B. T., Carriedo, L., Maloof, J. N., & Sinha, N. (2016). In-field plant phenotyping using multi-view reconstruction: An investigation in eggplant. *Proc. 13th Intl. Conf. Precision Agriculture*. Monticello, IL: International Society of Precision Agriculture.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybernetics*, 9(1), 62-66. <https://doi.org/10.1109/TSMC.1979.4310076>
- Perich, G., Hund, A., Anderegg, J., Roth, L., Boer, M. P., Walter, A., ... Aasen, H. (2020). Assessment of multi-image unmanned aerial vehicle based high-throughput field phenotyping of canopy temperature. *Front. Plant Sci.*, 11, article 150. <https://doi.org/10.3389/fpls.2020.00150>
- Salas Fernandez, M. G., Bao, Y., Tang, L., & Schnable, P. S. (2017). A high-throughput, field-based phenotyping technology for tall biomass crops. *Plant Physiol.*, 174(4), 2008-2022. <https://doi.org/10.1104/pp.17.00707>
- 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), article 214. <https://doi.org/10.3390/s17010214>
- Sodhi, P., Vijayarangan, S., & Wettergreen, D. (2017). In-field segmentation and identification of plant structures using 3D imaging. *Proc. IEEE/RSJ Intl. Conf. Intelligent Robots and Systems (IROS)* (pp. 5180-5187). <https://doi.org/10.1109/IROS.2017.8206407>
- Sun, S., Li, C., Chee, P. W., Paterson, A. H., Jiang, Y., Xu, R., ... Shehzad, T. (2020). Three-dimensional photogrammetric mapping of cotton bolls *in situ* based on point cloud segmentation and clustering. *ISPRS J. Photogram. Remote Sensing*, 160, 195-207. <https://doi.org/10.1016/j.isprsjprs.2019.12.011>
- Tu, S., Pang, J., Liu, H., Zhuang, N., Chen, Y., Zheng, C., ... Xue, Y. (2020). Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images. *Prec. Agric.*, 21(5), 1072-1091. <https://doi.org/10.1007/s11119-020-09709-3>
- Wu, S., Wen, W., Wang, Y., Fan, J., Wang, C., Gou, W., & Guo, X. (2020). MVS-Pheno: A portable and low-cost phenotyping platform for maize shoots using multiview stereo 3D reconstruction. *Plant Phenomics*, 2020, article 1848437. <https://doi.org/10.34133/2020/1848437>
- Xin, Z., Li Wang, M., Barkley, N. A., Burow, G., Franks, C., Pederson, G., & Burke, J. (2008). Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. *BMC Plant Biol.*, 8(1), article 103. <https://doi.org/10.1186/1471-2229-8-103>
- Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-R-CNN. *Comput. Electron. Agric.*, 163, article 104846. <https://doi.org/10.1016/j.compag.2019.06.001>
- Zhang, D., Dai, L., Xiang, S., & Zhang, D. (2015). Segment graph based image filtering: Fast structure-preserving smoothing. *Proc. IEEE Intl. Conf. Computer Vision* (pp. 361-369). <https://doi.org/10.1109/ICCV.2015.49>

Author

First Name	Middle	Surname	Suffix	Role	E-mail	Contact author?
Lirong		Xiang			xiang@iastate.edu	No

Affiliation

Organization	Address	Country	URL
Department of Agricultural and Biosystems Engineering	Iowa State University, Ames, Iowa	USA	

Author

First Name	Middle	Surname	Suffix	Role	E-mail	Contact author?
Lie		Tang			lietang@iastate.edu	yes

Affiliation

Organization	Address	Country	URL
Department of Agricultural and Biosystems Engineering	Iowa State University, Ames, Iowa	USA	

Author

First Name	Middle	Surname	Suffix	Role	E-mail	Contact author?
Jingyao		Gai			jygai@iastate.edu	No

Affiliation

Organization	Address	Country	URL
Department of Agricultural and Biosystems Engineering	Iowa State University, Ames, Iowa	USA	

Author

First Name	Middle	Surname	Suffix	Role	E-mail	Contact author?
Le		Wang			lewang@iastate.edu	No

Affiliation

Organization	Address	Country	URL
Department of Agricultural and Biosystems Engineering	Iowa State University, Ames, Iowa	USA	

Publications Staff to Complete

Start Date (generally one year from publication date)				End Date (blank if none)		
Copyright?	Year	Journal	Volume, Issue	DOI	Manuscript ID	
Yes	2021	Transactions of the ASABE	64(5)	10.13031/trans.14156	ITSC14156	