2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation
DOI: https://doi.org/10.13031/aim.202001190
Paper Number: 2001190

PhenoStereo: a high-throughput stereo vision system for field-based plant phenotyping - with an application in sorghum stem diameter estimation

Lirong Xiang[1], Lie Tang[1,2*], Jingyao Gai[1], Le Wang[1,3]

1 Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011, USA

2 Plant Sciences Institutes, Iowa State University, Ames, IA 50011, USA

3 Department of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, 310058, China

**Correspondence: lietang@iastate.edu*

Written for presentation at the
2020 ASABE Annual International Meeting
Sponsored by ASABE
Omaha, Nebraska
July 12–15, 2020

ABSTRACT. ***In recent years, three-dimensional (3D) sensing has gained a great interest in plant phenotyping because it can represent the 3D nature of plant architecture. Among all available 3D imaging technologies, stereo vision offers a viable solution due to its high spatial resolution and wide selection of camera modules. However, the performance of in-field stereo imaging for plant phenotyping has been adversely affected by textureless regions and occlusions of plants, and variable outdoor lighting and wind conditions. In this research, a portable stereo imaging module namely PhenoStereo was developed for high-throughput field-based plant phenotyping. PhenoStereo featured a self-contained embedded design, which made it capable of capturing images at 14 stereoscopic frames per second. In addition, a set of customized strobe lights was integrated to overcome lighting variations and enable the use of high shutter speed to overcome motion blurs. The stem diameter of sorghum plants is an important trait for stalk strength and biomass potential evaluation but has been identified as a challenging sensing task to automated in the field due to the complexity of the imaging object and the environment. To that connection, PhenoStereo was used to acquire a set of sorghum plant images and an automated point cloud data processing pipeline was also developed to automatically extract the stems and then quantify their diameters via an optimized 3D modeling process. The pipeline employed a Mask R-CNN deep learning network for detecting stalk contours and a Semi-Global Block Matching 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 approaches. These results demonstrated that with proper customization stereo vision can be a highly desirable sensing method for field-based plant phenotyping using high-fidelity 3D models reconstructed from stereoscopic images. With the proving results from sorghum plant stem diameter sensing, this proposed stereo sensing approach can likely be extended to characterize a broad spectrum of plant phenotypes such as leaf angle and tassel shape of maize plants and seed pods and stem nodes of soybean plants.***

Keywords. field-based high-throughput phenotyping, point cloud, stem diameter, stereo vision.

Introduction

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). Applications of genomic technologies, however, are limited by the ability of accurate and precise phenotypic trait collection. Traditional phenotyping procedures are time-consuming, labor-intensive, and low-throughput, whereas most phenotypes were obtained in a destructive way or involved manual observations (Minervini et al., 2017). Hence, the efficient and effective phenotyping strategies are demanded 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 major categories: aerial-based and ground-based measuring systems (Li et al., 2014). Aerial-based phenotyping platforms enable capturing information of a large scale 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 index (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, 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 sensors payload, and high sensor resolution. With customized ground vehicles and a wide range of sensing 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 ‘Robotanist’, ‘Vinobot’, and some semi-autonomous tractor-based systems (Milella et al., 2019; Salas Fernandez et al., 2017b). 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 like fluorescence, thermal, hyperspectral, and three-dimensional (3D) sensors have been integrated into phenotyping systems. Among these sensing technologies, 3D sensors have gained a great interest especially for measuring architectural features because of the 3D nature of plant architecture. Light detection and ranging (LiDAR) sensor is a widely used 3D sensor for field-based phenotyping. Chakraborty et al. (2019) reconstructed 3D models of apple trees and grapevine canopies by combining a 3D LiDAR with an inertial measurement unit (IMU) on a ground vehicle. The canopy volume was derived from the 3D models and found to be strongly correlated with manual measurements. Besides canopy-level morphological traits, high-resolution 3D LiDAR sensors have also provided the potential for organ-level traits for organs such as cotton balls (Sun et al., 2020a), sorghum panicles (Malambo et al., 2019), and corn stems (Jin et al., 2019). However, such measurements can be complex and require longer imaging time to obtain dense and accurate canopy models (Dandrifosse et al., 2020). Time-of-Flight (ToF) camera is another widespread sensor which 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 (Ji Li and Tang, 2018). The major weakness of ToF sensors is its sensitiveness to strong sunlight. To alleviate this issue, a shroud or an umbrella could be used to reduce the sunlight intensity in outdoor applications (Gai et al., 2020; Mortensen et al., 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 utilized 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 can be estimated from the system. MVS-SFM method is capable of generating dense point cloud of plants but is challenging to implement in the field, especially for tall-growing plants like corn/sorghum. Binocular stereo vision is being considered as an alternative option to overcome limitations with MVS-SFM technique. The technique only needs two cameras to compute depth information. Previous work has demonstrated the potential of stereo vision in measuring plant architecture outdoors. For example, Sodhi et al. (2017) implemented such a system to study sorghum plant architecture in field conditions. In their application, phenotypic traits like 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, where the stereo cameras were positioned laterally and stacked vertically on an extension rig to perform side-view imaging. Several morphological traits were quantified and found to be accurate and highly repeatable. However, the wider

row spacing was required by the system. Though 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 to drought/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., 2017b; Zhao et al., 2016). Typically, sorghum stem diameter is measured manually using calipers - a tedious and error-prone practice, but to automate a process that can accurately measure the stem diameter of sorghum plants in the field has been a challenging task (Salas Fernandez et al., 2017a). 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 retrieve stem diameter from side-view images, where a user would first click four reference points on the stem edge to identify a representative stem, the stem diameter was then calculated by the triangulation principle. Though results showed high repeatability and accuracy, the method requires human intervention. Recent advances in machine learning offered new approaches for reliable object detection in some image-based deep learning tasks (Gené-Mola et al., 2020; J. Jin et al., 2018). For example, Baweja et al. (2018) used convolutional deep neural networks to detect sorghum stems and calculated 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 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., 2020b).

Based on the aforementioned literature research, this research project was set out to investigate a robust and accurate machine vision system that can measure the stem diameter of sorghum plants in 3D space under outdoor field conditions. The developed vision system was expected to deliver consistent performance under 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) develop a customized stereo module for high-quality and high-throughput image acquisition in the field and between the agronomically spaced crop rows; 2) develop an automated image processing pipeline to detect the stems of sorghum plants and measure their stem diameter in reconstructed 3D model; 3) evaluate the performance of stem instance segmentation and stem diameter estimation.

Material and Method

Development of a stereo imaging system

PhenoStereo, which refers to a custom-built stereo imaging module for plant phenotyping, is capable of capturing high-quality images in the field with strobe lighting. The PhenoStereo (fig. 1a) 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, California, USA), an Elroy carrier board (Connect Tech Inc., Ontario, Canada), and other devices (e.g, SD card, PCIe card) on the carrier board. Jetson is able to simultaneously take image pairs from two stereo cameras through high-speed interfaces (PCIe) and save them to a SD card. The Elroy Carrier board was chosen to interface with the Jetson TX2 module due to its small size and a variety of standard hardware interfaces.

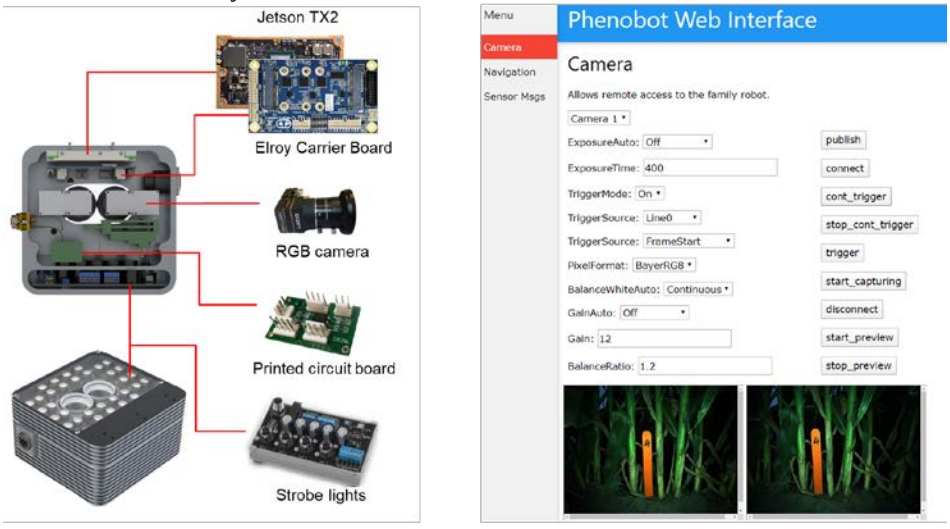

Figure 1. The complete device. (a) Configuration of PhenoStereo imaging unit: consisting of stereo cameras and strobe lights for capturing high-resolution images. (b) The user interface of PhenoStereo, which uses Robot Operating System (ROS) network for wireless control.

Two identical RGB cameras (Phoenix 3.2MP, Lucid Vision labs, Canada) equipped with a lens of 4.0 mm focal length were used. The baseline of the stereo pair was set at 38 mm, which enables the module to image close by objects with a large overlap area between the left and right images. The stereo camera has a horizontal view angle of 85.8° and a vertical view angle of 63.6°. The 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, were 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, USA) were mounted around the camera pair for strobing illumination. The motivation behind this design is to provide consistent lighting, enhance 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, USA) for driving high-intensity LEDs.

The imaging platform was enclosed in a waterproof metal case (151 *mm* × 146 *mm* × 82 *mm*), and the system weighs approximately 2.06 kg. An electronic circuit was designed to synchronize and trigger the camera pair and the strobe lights. The imaging system was able to acquire images with light illumination of 31000 lux at 0.5 m. By using the Robot Operating System (ROS) JavaScript Library, a web-based user interface (fig. 1b) was developed to control the cameras and visualize the images. The interface allows a user to adjust camera parameters (exposure time, white balance, etc.), send trigger commands, and view live images on a smartphone/laptop browser. The live images and commands are published over local area network (LAN) using ROS. The control algorithm for image acquisition was written in C++, using ROS Kinetic running on Ubuntu 16.04.

Data collection

A 4-wheel custom-built ground vehicle was used to deploy the PhenoStereo (fig. 2). The PhenoStereo was mounted on the vehicle at a height of 20 inches above the ground, and aluminum frames were used to support the camera module. The ground vehicle was driven at 3 mph (4.8 km/h) through 30-inch spaced crop rows and the images were captured at 10 FPS. To prevent motion blur caused by robot movement and wind conditions, the shutter speed was set to 0.3 ms. The distance between the camera and the sorghum stalks was around 15 inches (0.38 m).

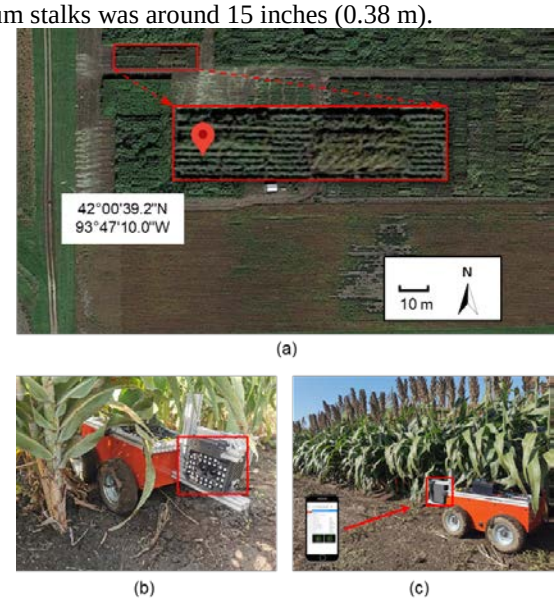

Figure 2. Data collection. (a) Location of the image collection. (b-c) Remote-controlled data collection with PhenoStereo under sunny (b) and overcast (c) weather conditions. A smartphone was used to interact with the PhenoStereo using wireless network connection.

Data were acquired at the Agricultural Engineering and Agronomy Research Farm of Iowa State University in Boone, Iowa (fig. 2a). The data collection was carried out on Sep. 30th and Oct. 9th, 2019. In order to test the proposed module's robustness against varying outdoor lighting conditions, sorghum plant images from the field under sunny (fig. 2b) and overcast (fig. 2c) weather conditions were obtained. 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 at the stem section of 5-10 cm above the soil surface on Oct. 9th. In total there were 75 stems measured to evaluate the performance of stem diameter estimation.

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 stereo images by Semi-Global Block Matching (SGBM) (Hirschmüller, 2008). Then, individual stem masks were detected using 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 3D point cloud to segment individual stems. After that, noisy voxels in the segmented point cloud were detected and filtered using local geometric features. At last, cylinder fitting was conducted on selected stem sections, the stem diameters were finally derived from the fitted cylinders. Additional details are described in the following sections.

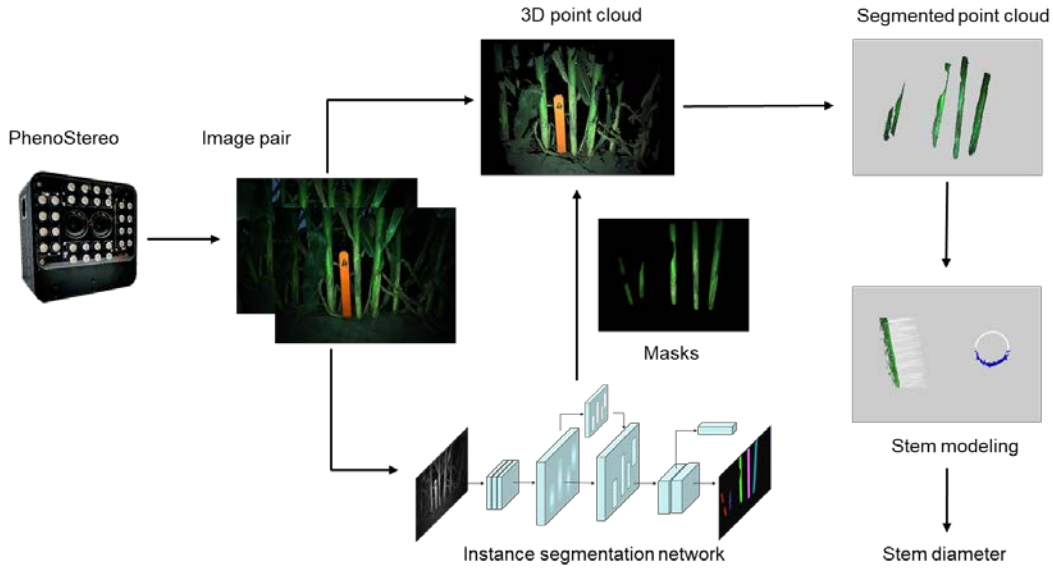

Figure 3. Illustration of the proposed image processing pipeline.

3D reconstruction

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

The camera calibration aims to compute the intrinsic and extrinsic parameters of stereo cameras. The camera intrinsic parameters include the focal length, the pixel size, and the principle point. The extrinsic parameters consist of a rotation and a translation matrix, which are related to the coordinate transformation between an object in 3D world and the 2D images. In our calibration procedure, a checkerboard with squares of size $20\ mm \times 20\ mm$ was used to obtain 30 image pairs. During the image acquisition process, the relative distances and the poses between the checkerboard and the camera were changed. All the corners of the squares on the checkerboard were detected to serve as feature points, the camera parameters were established by calculating the relationship between the feature point sets and their corresponding 3D world coordinates (Bradski and Kaehler, 2008). With the camera intrinsic parameters, the image pair can be rectified by applying a projective transformation. Since the images follow the principle of epipolar constrains (Brown et al., 2003), two corresponding points are on the same horizontal line in the rectified image pair (fig. 4a). The rectification process removes the lens distortion and enables the reduction of the search space of stereo correspondence from 2D to 1D.

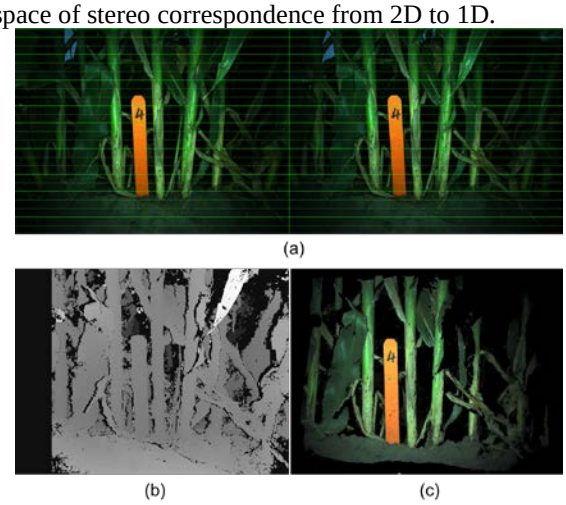

Figure 4. The procedure of stereo matching using Semi-Global Block Matching (SGBM). (a) A rectified and segmented image pair, where corresponding points are on the same horizontal scan lines. (b) The disparity output of the SGBM algorithm. (c) 3D point cloud reconstruction of the sorghum plant.

To extract the 3D coordinates of the object, the rectified image pair was used to generate a disparity map by stereo matching. Stereo matching is the process of finding corresponding pixels in a pair of stereo images. For stereo correspondence matching of field crops, two thorny situations can happen. The first one is from the nearly homogeneous color and texture of plant canopy surfaces that can make the correspondence search ambiguous. The second challenging situation is occlusion, which means that some pixels in one image do not have correspondences in the other image. In this study, the Semi-Global Block Matching (SGBM) (Hirschmüller, 2008) was adopted to do the stereo matching given its efficiency and robustness. 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 constrain. With global consistency constraints in the optimization process, the SGBM has shown the ability to retain edges and deal with untextured areas (Hirschmüller, 2008). Considering the PhenoStereo has a short baseline (38 mm) and the sorghum plants were around 15 inches away from the camera, the matching window size and the disparity range were set to 5 pixels and 100 pixels, respectively. The result of stereo matching is the disparity map (fig. 4b), which gives the apparent pixel difference in left and right images. The disparities are inversely proportional to 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, 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 represent the coordinates of a pixel in the 2D image, and d is the corresponding disparity value, b denotes the baseline (m) of the stereo camera, and f is the focal length (pixel). The focal length and baseline can be obtained from the calibration process. After the stereo matching, a 3D point cloud can be obtained (fig. 4c). Since we are only interested in plants with a distance of 15-20 inches, the points with z-coordinates larger than 0.8 m or less than 0.002 m were removed as background/noise. The obtained 3D model is processed to determine the stem diameter according to geometric features.

Stem segmentation

For individual stalk size measurement, effective stem segmentation is the prerequisite for further image analysis. Though there was a strong contrast between plants and background in the images, identifying the stem from plant canopy can be a difficult task because of the color and texture similarities between plant organs. Additionally, the occlusions generated by the dense canopy of sorghum plants have proposed challenges for automated segmentation of stems. In this study, we firstly detect the individual stems in 2D images, the 2D detections were then rectified and projected onto the 3D point cloud to segment the stems.

The Mask R-CNN, which provides both bounding boxes and semantic masks for the detected object regions, was utilized to detect and segment sorghum stems from RGB images (fig. 5a). The model is an improvement over Faster R-CNN (Ren et al., 2017), where a mask generating branch was added to predict segmentation masks on each output proposal box. The framework of Mask R-CNN (fig. 5b) consists of two parts: 1) a Convolutional backbone part for feature map extraction from input images; and 2) a head part which takes the feature maps for classification prediction, bounding box recognition, and mask prediction. The backbone employs a feature pyramid network (FPN) (Lin et al., 2017) to construct feature maps for objects at different scales. The feature maps are then fed 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. The RPN slides different region boxes (anchors) on the feature map to generate ROIs that are most likely to contain objects. In this study, according to the elongated shape of the target (sorghum stems), three length-width ratios (2:1, 4:1, and 8:1) of the anchors were used to improve computational efficiency.

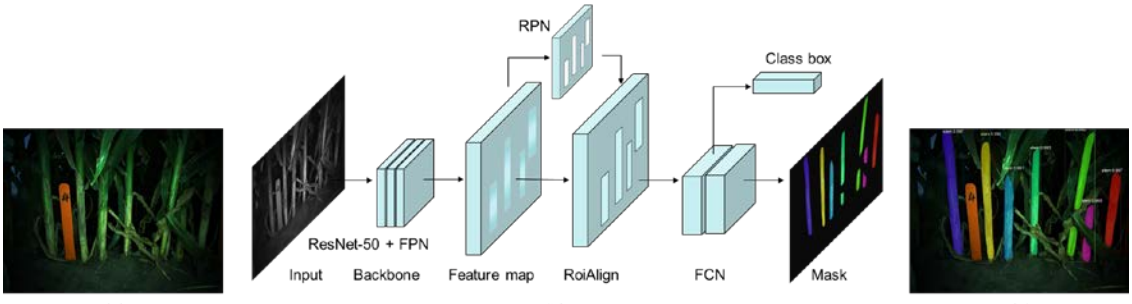

Figure 5. Summary of Mask R-CNN based stem segmentation. (a) An example of the input RGB image. (b) The diagram of Mask R-CNN architecture. (c) A masked image with detected stems.

The detection of individual stems was formulated as a semantic segmentation task with two classes: 'stem' and 'background'. Pixels of the class 'stem' referred to the stalk of sorghum plants, while remaining pixels were assigned to the class 'background'. In this case, the background not only included field background, but also other non-stem parts of the plant, such as leaves and tillers. We manually annotated 440 images, where each plant stem was colored independently. The labeled images were split into a training set, a test set, and a validation set with a proportion of 7:2:1. Although the strobe lights were utilized to maintain consistent lighting conditions, 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 dimension from three channels to one channel, which also decreases the complexity of network architecture.

We used a TensorFlow-based implementation of the Mask R-CNN with a ResNet-50 backbone network. The Mask R-CNN model was initialized using pre-trained weights from COCO dataset (Lin et al., 2014). Image augmentation was employed to increase the number of training images and diminish overfitting. The image augmentation technique allows the networks to be more adaptive to field conditions. In this study, 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). During the training process, the augmentations were randomly assigned to 50% of the source images to enlarge the training dataset artificially. The obtained masks were applied to the original image for the segmentation of individual stems (fig. 5c).

Point cloud filtering

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

During the filtering process, we denoted the stem points and non-stem points as inliers and outliers, respectively. It was noticed that most of the outliers were from leaf collar (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 stems' cylindrical structures, 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 enables the fast lookup of range and nearest neighbors. Principle 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. 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 $\lambda_1, \lambda_2, \lambda_3$ ($\lambda_1 > \lambda_2 > \lambda_3$).

$$\mathcal{N}_k = \{p_i : \|p - p_i\| < d\} \tag{2}$$

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

$$\bar{p} = \frac{1}{k}\sum_{i=1}^{k} p_i \tag{4}$$

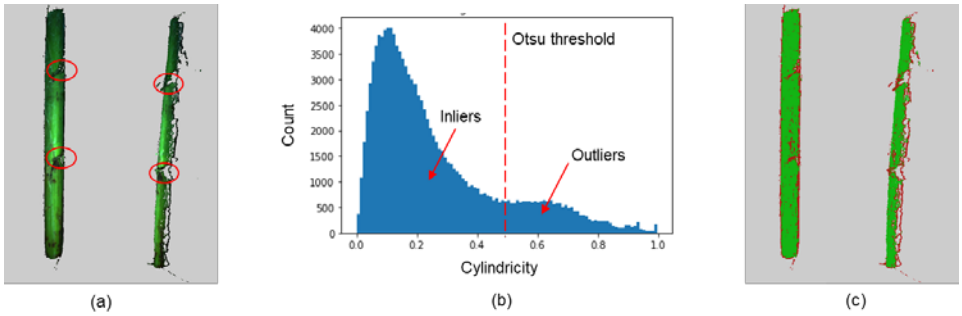

Figure 6. Point cloud filtering using local features. (a) The point cloud segment of a stem, where leaf collars were highlighted with red circles. (b) Histogram of cylindrical difference in stem and non-stem points. The red dotted line is the Otsu threshold which separates stem inliers from noise. (c)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

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 isotropic spatial distribution, such as a spherical structure. For a cylindrical structure like sorghum stem, we expect λ_3 to be much smaller than λ_1 and λ_2 ($\lambda_1 > \lambda_2 >> \lambda_3$). In this work, *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. Numerically, a point and its neighbor points belong to a sorghum stem should have small c values. The cylindricity value of each point was chosen to eliminate leaf collar 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), the points with cylindricity values greater than the Otsu threshold were removed as outliers (fig. 6c).

$$c = \frac{\lambda_1 - \lambda_2}{\lambda_1 - \lambda_3} \tag{5}$$

Stem diameter estimation

Fitting a cylindrical shape of 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. A cylinder is described by its axis and radius. If a point locates at a distance r from the axis, then it is on the cylindrical surface. Cylinder fitting can be a computationally intensive process because of its high parametric space. To reduce the complexity/dimension of the parameter space, we split the process to firstly find the cylindrical axis direction and then detect a circle in a plane. In the first step, the direction of the cylindrical axis was calculated from the surface normal. After that, the point cloud of a selected stem section was projected onto a 2D plane along the direction of the axis; a circle was then fitted to the projected points to find the radius, which 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 perpendicular to the tangent plane at P (fig. 7a). It's not hard to find 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 is computed by finding the eigenvector corresponding to λ_3, which is the smallest eigenvalue of the covariance matrix in equation 3. We calculated surface normal for each individual point in the point cloud (fig. 7b and c). We aim to find a vector that is most orthogonal to the surface normal 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.

$$\mathcal{F}(a,b) = \sum_{i=1}^{m} (\vec{\omega} \cdot \overrightarrow{n_i})^2 = \sum_{i=1}^{m} (a \cdot x_i + b \cdot y_i + z_i)^2 \tag{6}$$

where $\overrightarrow{n_i}(x_i,\ y_i, z_i)$ is the normal vector of the i[th] point in the point cloud, and *m* is the number of points.

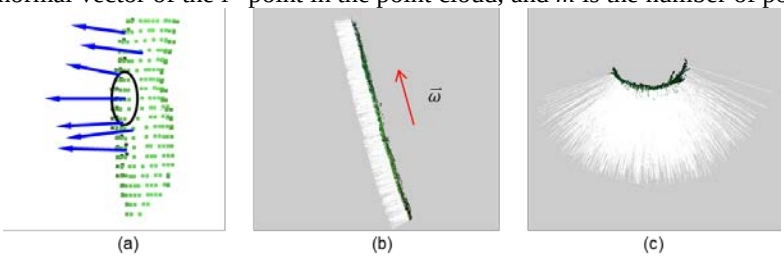

Figure 7. The estimation of the direction of the cylindrical axis. (a) Illustration of surface normal estimation. (b) The side view of the surface normals of a stem, the vector $\vec{\omega}$ represents the direction of the cylindrical axis and the white epidermal hairs are the surface normals. (c) The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

Before projecting the point cloud to a 2D plane, the stem was rotated such that the cylindrical axis is aligned with z coordinates. The direction vector of the cylinder axis $\vec{\omega}$ was normalized as $\overrightarrow{\omega_0}$. The unit vector $\overrightarrow{u_z}(0, 0, 1)$ 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 $\overrightarrow{\omega_0}$ to vector $\overrightarrow{u_z}(0, 0, 1)$ can be obtained.

$$\vec{v} = \overrightarrow{\omega_0} \times \overrightarrow{u_z} \tag{7}$$

$$k = \overrightarrow{\omega_0} \cdot \overrightarrow{u_z} \tag{8}$$

$$R\ = I + [\sim\vec{v}] + [\sim\vec{v}]^2 \cdot \frac{1}{1+k} \tag{9}$$

where I is the identity matrix, and $[\sim\vec{v}]$ (eq. 10) is the skew-symmetric matrix of $\vec{v}$.

$$[\sim\vec{v}] = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix} \tag{10}$$

Finally, the stem point cloud was rotated by matrix *R* to ensure that the cylindrical axis was parallel to z-axis (fig. 8a). After rotation, the points with z values from -15 to -10 cm, which approximately represented the stem section of 5-10 cm above the ground, were selected for stem diameter estimation. The selected point cloud was projected to x-y plane, where

the projected points distributed as a partial circle (fig. 8b). The center of the circle (x_0, y_0) and the radius R were optimized by performing an unconstrained minimization of $\mathcal{F}(x_0, y_0, R)$ in the three-dimensional parameter space (eq. 11).

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

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

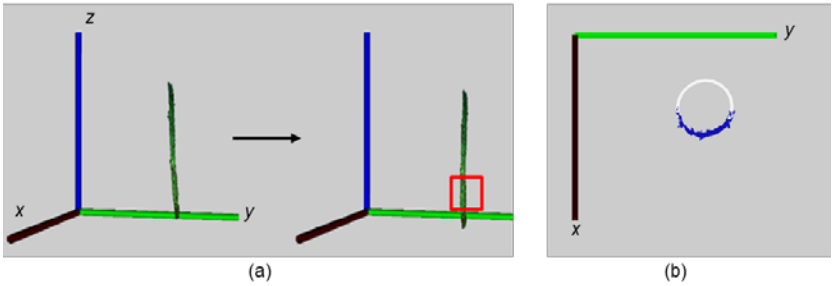

Figure 8. Circle detection using Levenberg-Marquardt algorithm (LMA). (a) 3D transformation of the stem point cloud, the point cloud was rotated to have its cylindrical axis to be parallel with z-axis. The stem section in the red square was selected for circle fitting. (b) Circle fitting in x-y plane,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

The Levenberg-Marquardt algorithm (LMA) (Levenberg, 1944; Marquardt, 1963) was used for the optimization process. At k[th] iteration, LMA uses the following equation to update:

$$x_{k+1} - x_k = ({J_k}^T J_k + \lambda_k I)^{-1} {J_k}^T \varepsilon_k \quad (12)$$

where J_x = Jacobian matrix, I = Identity matrix, and ε_k is a vector of errors. LMA interpolates between the Gauss-Newton (GN) method and the steepest-decent (SD) method through the adjustment of λ_k (Song et al., 2020). For example, when λ_k is assigned a small value, the descent is close to GN method, while a large λ_k value makes the LMA step close to the SD method.

Accuracy assessment

To evaluate the performance of stem segmentation, three widely used evaluation metrics of precision (P), recall (R), and harmonic mean (F_1) were chosen, 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 (True Positives) denotes the number of correctly segmented stem pixels; FP (False Positives) is the number of background pixels incorrectly classified as stem pixels; and FN (False Negatives) is the number of non-segmented stem pixels.

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).

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

$$MAE = \frac{1}{N} \sum_{i}^{N} \left| d_i - d_i^{gt} \right| \quad (17)$$

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

Table 1. The results of instance segmentation by Mask R-CNN

IoU	P	R	F1
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

The test dataset was employed to further evaluated 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 reported precision, recall, and F_1 over a variety of Intersection-over-Union (IoU) thresholds from 0.4 to 0.7 (Table 1). The IoU indicates the overlap rate between the predicted mask and the ground truth. Considering IoU values equal to or lower than 0.5, all the precision, recall, and F_1 rates are higher than 0.8, which can meet the need of stem instance segmentation in this study. Furthermore, the precision, which represents the ratio of correctly classified pixels in all returned pixels, is impressive even considering a 0.7 IoU. For each IoU level, it is obvious that the recall rate was lower than precision rate. This is mainly because of the non-detected stem pixels close to the soil, where the texture and color were not prominent enough. However, since we only measured the diameter of the stem section from 5-10 cm above the ground, the misdetection of the pixels near the soil will not affect the stem diameter estimation performance.

The proposed Mask R-CNN segmentation approach produced satisfactory results for instance segmentation. It can be seen that the stem mask extraction is very close to the real stem contour (fig. 9). For each segmented instance, the confidence level was close to 1. Also, the high statistic values (P, R, and F1) demonstrated the robustness of the trained model. Despite the relatively small size of training images, the model is adaptive to variable texture, stem size, occlusion, and illumination conditions. For example, the stem with a yellow mask in the first column (fig. 9) was visually split by a leaf. With the trained model, the target stem was successfully recognized as one single domain with two separate regions. Moreover, it can be seen that the proposed approach is not influenced by the reduction of lighting intensity near image borders.

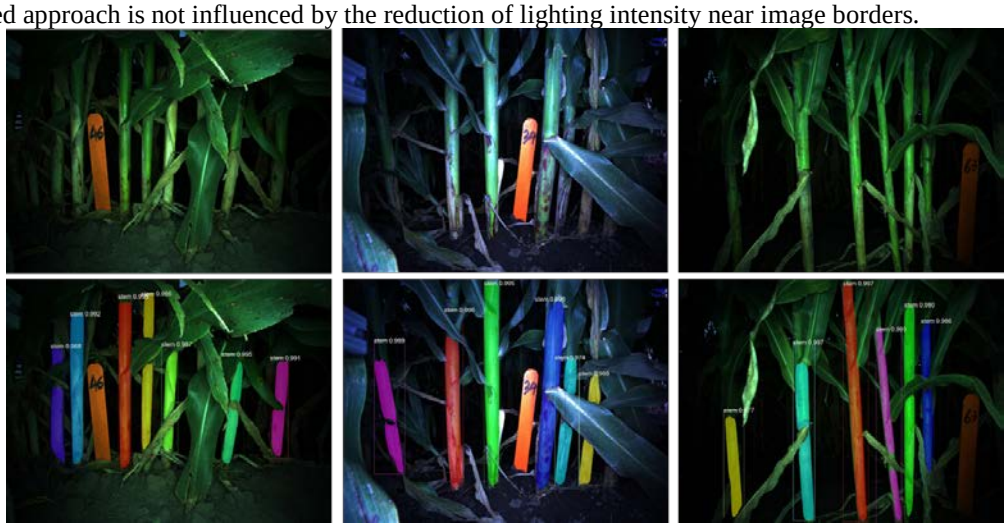

Figure 9. Some instance segmentation results produced by Mask R-CNN. Examples of the original images are displayed in the top row and their corresponding segmentation results are illustrated in the bottom row. Each color area indicates a segmented stem instance.

Performance of stem diameter estimation

The image-derived stem diameters were found to be highly correlated (r = 0.97) and precise (RMSE = 1.39 mm) with respect to manually measured stem diameters (fig. 10). Errors in system-derived diameters were attributable to several factors including falsely detected stem edges, inaccurate stereo matching, and camera calibration imperfection. The calibration errors resulted in an inaccurate estimation of the extrinsic as well as the 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 image-derived diameter was slightly larger than the corresponding ground truth. This is mainly caused by the soft leaf sheath around the stems, as the caliper tended to give a smaller value when tightening across the stems.

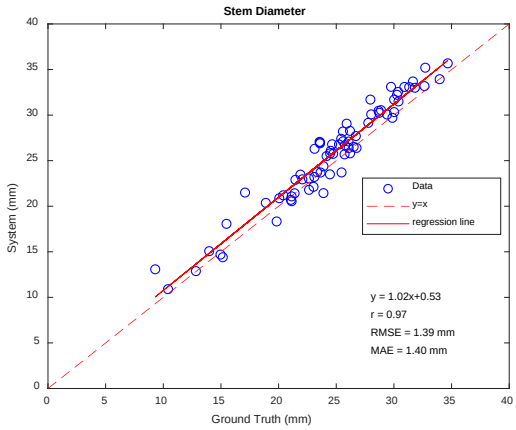

Figure 10. Correlation of system-derived stem diameter and ground truth

It was noticed that the proposed approach showed lower reliability when the stem diameter was less than 20 mm. One possible explanation is the inherent difficulty to reconstruct 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 was small, therefore, the process of stem modeling and the placement of caliper can introduce errors.

Overall, the proposed approach outperformed the current automated and user-interactive methods. The automated method of StalkNet (Baweja et al., 2018) showed promising efficiency in the measurement of stem width, but with the 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 mm) with our automated pipeline, but the intercept (2.84 mm) was substantially larger than that of our method. Additionally, their method requires human intervention to assist in the segmentation of stems, which is still laborious and not practical 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 of using high-accuracy 3D surface models to estimate the stem dimeter 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, 32GB RAM, and a NVIDIA Titan Xp GPU. The instance segmentation network was trained on a Nvidia GTX Titan X GPU with a 3.5 GHz Xeon HexaCore CPU and 16 GB RAM. We summarized the time cost of each procedure in Table 2. Here the time for image processing consists of two parts: data extraction and network training. The data extraction comprised four processing stages: stereo matching, stem detection, point cloud filtering, and stem diameter estimation. Usually, the time 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, which occupies 49% of the total processing time. During the filtering process of a dense point cloud, traversing each point to search for its neighbors costs much time. However, the proposed stem diameter extraction operating on the point cloud was fast, and the average processing time was 3.8 s per stem. After evaluated 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 that of in-field manual measurement. The Mask R-CNN training process lasted 35 hours with a 0.001 learning rate.

Table 2. The average computation time for each processing stage

Processing stage	Average time (s)
Stereo matching	5.4s per image
Stem detection	1.4s per image
Point cloud filtering	10.3s per stem
Stem diameter estimation	3.8s per stem
Mask R-CNN training	35 hours

Discussion

Performance of PhenoStereo

In field conditions, the development of an imaging system that is robust to environmental conditions is challenging. Herein, we discussed three major challenges that come with field-based imaging and our corresponding solutions, the challenges include: 1) variable illumination, 2) wind, and 3) complex background. Specifically, 1) In the field, illumination is a key factor affecting the quality of images. The direct sunlight and backlighting conditions can lead to over- or under-exposed images. Such an illumination issue can also reduce image texture and hence affect the performance of stereo reconstruction. The stereo matching requires that the plant images have as rich and crisp texture as possible. The PhenoStereo

comprises strobe lights which actively illuminate the object when capturing images, thus increasing the local discriminability and sharpness of the images; 2) The 3D reconstruction can be adversely affected by blurry images induced by wind conditions. In this study, the high-powered flash combined with extremely short exposure time enabled the camera module to capture images in a short time (0.3 ms), which effectively reduced the adverse effects of motion due to wind. Meanwhile, the fast shutter speed also decreases the influence of robot motion and vibrations on image quality; 3) The complex backgrounds in the field have imposed steep challenges for image segmentation and stereo reconstruction. The proposed PhenoStereo, with dominating lighting on the viewed object, was shown to be capable of producing images with a strong contrast between the foreground and background.

Field-based phenotyping has been recognized as an important alternative for UAV-based phenotyping, for its capability of delivering organ-level phenotypic traits located at the middle and bottom of the plant canopy. However, most current stereo-vision-based phenotyping platforms have been reported to work best under uniform lighting condition and with the absence of wind (Dandrifosse et al., 2020; Kaczmarek, 2017; Wu et al., 2020). Nevertheless, the PhenoStereo has been successfully employed to collect over 10,000 image pairs of maize and sorghum, under various weather conditions and at different timings of day. The consistent and high-quality images demonstrated the robustness and generality of the PhenoStereo camera module. Moreover, the PhenoStereo is equipped with ROS-based API and Gigabit Ethernet Interface, making it highly integrative with many field-based phenotyping systems. Compared with manual imaging, the high data acquisition speed efficiency makes the module feasible to acquire phenotypes at large scales.

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 lens to reduce noise caused by glares. The stem and leave surfaces of plants tend to be reflective, especially when the camera getting too close to the viewed object, result in a negative effect on image quality (fig. 11). Another possibility would be to use more than one camera set to build a stereo vision system. By taking advantage of multiple adjacent cameras, the system can have a wider field of view, allowing some side cameras to have advantageous view angles over areas containing partial objects. The multi-camera system offers the potential to generate more accurate distance estimations and better view angles over occluded areas (Kaczmarek, 2017).

Figure 11. The partially over-exposed images due to the reflective surface of sorghum stems.

Concerning the image processing pipeline, even more advanced stereo matching algorithms could be used to improve the quality of disparity maps. This study focuses on SGBM for computing stereo correspondence due to its high efficiency and sufficient performance. However, it is worth to point out that SGBM has its limitations. In this study, the surface reconstructed from SGBM was found to be blurry near the object boundaries, which gave inaccurate contours of stem and leaf. This is because the SGBM's filter-based matching cost aggregation tend 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-preserving performance, including the optimization based filters and the 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 canopy in the field. The algorithm takes the advantages of both CNNs for computing matching cost and nonlocal support region filtering for preserving edges. Though with lower computational efficiency, the 3DMST outperformed SGBM in surface smoothness and edge-preserving ability regarding the surface reconstruction of sorghum plants.

Conclusion

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 facilitated the application of the PhenoStereo under various environmental conditions (direct sunlight, backlighting conditions, shadows, wind conditions). As a case study, we used the PhenoStereo to quantify the stem diameter of sorghum plants. The high-quality stereo images allowed the reconstruction of accurate surface models of plant stems. Subsequently, an automated image processing pipeline was developed to segment individual stems and performed

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 diameter of sorghum plants. Our study also demonstrated that with proper customization stereo vision is feasible for 3D-based plant phenotyping under field conditions. Future work will focus on improving the quality of disparity maps, as well as using the PhenoStereo to characterize other organ-level morphological traits such as leaf angle and panicle/tassel size.

Acknowledgments

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 Panelo for his contributions to the collection of ground-truth data.

Reference

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. *Journal of Field Robotics*, *36*(2), 397–415. https://doi.org/10.1002/rob.21830

Bao, Y., Tang, L., Srinivasan, S., Schnable, P. S. (2018). ScienceDirect Field-based architectural traits characterisation of maize plant using time-of-flight 3D imaging. *Biosystems Engineering*, *178*, 86–101. https://doi.org/10.1016/j.biosystemseng.2018.11.005

Barbedo, J. G. A. (2019). A Review on the Use of Unmanned Aerial Vehicles and Imaging Sensors for Monitoring and Assessing Plant Stresses. *Drones*, *3*(2), 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, (c), 271–284. 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.). O'Reilly.

Brown, M. Z., Burschka, D., Hager, G. D. (2003, August). Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 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. *Computers and Electronics in Agriculture*, *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. *Computers and Electronics in Agriculture*, *158*, 284–293. https://doi.org/10.1016/j.compag.2019.02.012

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. *Frontiers in Plant Science*, *11*(February), 1–15. 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. *Journal of Intelligent and Robotic Systems: Theory and Applications*, *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 in Plant Science*, *16*(12), 635–644. https://doi.org/10.1016/J.TPLANTS.2011.09.005

Gai, J., Tang, L., Steward, B. L. (2020). Automated crop plant detection based on the fusion of color and depth images for robotic weed control. *Journal of Field Robotics*. https://doi.org/10.1002/rob.21897

Gené-Mola, J., Sanz-Cortiella, R., Rosell-Polo, J. R., Morros, J. R., Ruiz-Hidalgo, J., Vilaplana, V., Gregorio, E. (2020). Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. *Computers and Electronics in Agriculture*, *169*(December 2019), 105165. https://doi.org/10.1016/j.compag.2019.105165

Hackel, T., Wegner, J. D., Schindler, K. (2016). Contour detection in unstructured 3D point clouds. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 1610–1618. https://doi.org/10.1109/CVPR.2016.178

He, K., Gkioxari, G., Dollár, P., Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 386–397. https://doi.org/10.1109/TPAMI.2018.2844175

Hirschmüller, H. (2008). Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2), 328–341. https://doi.org/10.1109/TPAMI.2007.1166

Hirschmüller, H., Scharstein, D. (2007). Evaluation of cost functions for stereo matching. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2007.383248

Jin, J., Kohavi, G., Ji, Z., Zakhor, A. (2018). Top down approach to height histogram estimation of biomass sorghum in the

field. *IS and T International Symposium on Electronic Imaging Science and Technology*. https://doi.org/10.2352/ISSN.2470-1173.2018.15.COIMG-228

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 Transactions on Geoscience and Remote Sensing*, *57*(3), 1336–1346. https://doi.org/10.1109/TGRS.2018.2866056

Kaczmarek, A. L. (2017). Stereo vision with Equal Baseline Multiple Camera Set (EBMCS) for obtaining depth maps of plants. *Computers and Electronics in Agriculture*, *135*, 23–37. https://doi.org/10.1016/j.compag.2016.11.022

Krogh, A., Bender, A., Whelan, B., Barbour, M. M., Sukkarieh, S., Karstoft, H., Gislum, R. (2018). Segmentation of lettuce in coloured 3D point clouds for fresh weight estimation. *Computers and Electronics in Agriculture*, *154*(September), 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. *Quarterly of Applied Mathematics*, *2*(2), 164–168.

Li, J., Tang, L. (2018). Crop recognition under weedy conditions based on 3D imaging for robotic weed control. *Journal of Field Robotics*, *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), 1–13. https://doi.org/10.1186/s13007-019-0508-7

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

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. https://doi.org/10.1364/ao.56.003411

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* (Vol. 2017-January, pp. 936–944). Institute of Electrical and Electronics Engineers Inc. 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. In *European conference on computer vision* (pp.740-755). Springer, Cham. https://doi.org/10.1007/978-3-319-10602-1_48

Lu, N., Zhou, J., Han, Z., Li, D., Cao, Q., Yao, X., … Cao, W. (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*, 1–16. 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 Journal of Photogrammetry and Remote Sensing*, *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. *Journal of the Society for Industrial and Applied Mathematics*, *11*(2), 431–441. https://doi.org/10.1137/0111030

Milella, A., Marani, R., Petitti, A., Reina, G. (2019). In-field high throughput grapevine phenotyping with a consumer-grade depth camera. *Computers and Electronics in Agriculture*, *156*(May 2018), 293–306. https://doi.org/10.1016/j.compag.2018.11.026

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, 204–216. https://doi.org/10.1111/tpj.13472

Mortensen, A. K., Bender, A., Whelan, B., Barbour, M. M., Sukkarieh, S., Karstoft, H., Gislum, R. (2018). Segmentation of lettuce in coloured 3D point clouds for fresh weight estimation. *Computers and Electronics in Agriculture*, *154*, 373–381. https://doi.org/10.1016/j.compag.2018.09.010

Mueller-Sim, T., Jenkins, M., Abel, J., Kantor, G. (2017). The Robotanist: A ground-based agricultural robot for high-throughput crop phenotyping. In *Proceedings - IEEE International Conference on Robotics and Automation* (pp. 3634–3639). Institute of Electrical and Electronics Engineers Inc. 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. *13th International Conference on Precision Agriculture*, (July), 1–16.

Otsu, N. (1979). THRESHOLD SELECTION METHOD FROM GRAY-LEVEL HISTOGRAMS. *IEEE Trans Syst Man Cybern*, *SMC-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. *Frontiers in Plant Science*, *11*(February), 1–17. https://doi.org/10.3389/fpls.2020.00150

Ren, S., He, K., Girshick, R., Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Salas Fernandez, M. G., Bao, Y., Tang, L., Schnable, P. S. (2017a). A High-Throughput, Field-Based Phenotyping

Technology for Tall Biomass Crops. *Plant Physiology*, *174*(4), 2008–2022. https://doi.org/10.1104/pp.17.00707

Salas Fernandez, M. G., Bao, Y., Tang, L., Schnable, P. S. (2017b). A high-throughput, field-based phenotyping technology for tall biomass crops. *Plant Physiology*, *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 (Switzerland)*, *17*(1), 1–23. https://doi.org/10.3390/s17010214

Sodhi, P., Vijayarangan, S., Wettergreen, D. (2017). In-field segmentation and identification of plant structures using 3D imaging. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 5180–5187). IEEE. https://doi.org/10.1109/IROS.2017.8206407

Song, Z., Yang, F., Schonfeld, P., Li, J., Pu, H. (2020). Heuristic Strategies of Modified Levenberg-Marquardt Algorithm for Fitting Transition Curves. *Journal of Surveying Engineering*, *146*(2). https://doi.org/10.1061/(ASCE)SU.1943-5428.0000307

Sun, S., Li, C., Chee, P. W., Paterson, A. H., Jiang, Y., Xu, R., … Shehzad, T. (2020a). Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering. *ISPRS Journal of Photogrammetry and Remote Sensing*, *160*(May 2019), 195–207. https://doi.org/10.1016/j.isprsjprs.2019.12.011

Sun, S., Li, C., Chee, P. W., Paterson, A. H., Jiang, Y., Xu, R., … Shehzad, T. (2020b). Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering. *ISPRS Journal of Photogrammetry and Remote Sensing*, *160*(December 2019), 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. *Precision Agriculture*, 1–20. 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*, 1–17. 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 Biology*, *8*(1), 103. https://doi.org/10.1186/1471-2229-8-103

Zhang, F., Dai, L., Xiang, S., Zhang, X. (2015). Segment Graph Based Image Filtering: Fast Structure-Preserving Smoothing. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 361-369).

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), plantgenome2015.06.0044. https://doi.org/10.3835/plantgenome2015.06.0044