High-throughput robotic plant phenotyping using 3D machine vision and deep neural networks

by

Lirong Xiang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Agricultural and Biosystems Engineering

Program of Study Committee:
Lie Tang, Major Professor
Stuart Birrell
Brian L. Steward
Patrick Schnable
Yanhai Yin

Iowa State University

Ames, Iowa

2022

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENT

At first, I want to thank Dr. Lie Tang and his family for their support in my research and my life. I am very fortunate to have Dr. Lie Tang as my advisor, who has outstanding records in agricultural robotics, automation, and machine vision. Dr. Tang has generously guided me to conduct innovative research during my PhD study. He was also a tremendous source of advice, encouragement, and appreciation over the years. Dr. Tang understood the importance of fostering my independent research ability and offered constructive feedback for my career development. My life as a graduate student at Iowa State University (ISU) would have never been better without the remarkable appreciation and support from Dr. Tang and his family.

Meanwhile, I would like to give thanks to my POS committee members, Dr. Stuart Birrell, Dr. Brian L. Steward, Dr. Patrick S. Schnable, and Dr. Yanhai Yin for their constructive suggestions in my research and the support for my PhD program. It’s my great pleasure to have Dr. Brian L. Steward as a mentor of the course in the Preparing Future Faculty (PFF) program. He kindly provided me with guidance and resources with his time and patience whenever I seek help from him. The discussions with Dr. Steward about various aspects of being a professor (research, teaching, service, professional activities, etc.) has enabled me to understand faculty life much better. I am also indebted to Dr. Stuart Birrell for providing valuable comments on my research, Dr. Patrick Schnable for helping broaden my view on robotic plant phenotyping, and Dr. Yanhai Yin for his appreciation, encouragement, and support. I also want to thank Dr. Jianming Yu for his support and suggestions in the leaf angle project and Dr. Justin Walley for his help in the RoAD project.

This is a great chance to express my sincere thanks to all my colleagues and friends for their great support for both my research and life at ISU. I would like to thank the previous and current colleagues in the Robotics and Automation Lab: Dr. Yin Bao, Dr. Jingyao Gai,

Taylor Tuel, Dr. Le Wang, Dr. Wenyu Zhang, Duan Liu, Xuan Liu, Aditya Raj, Jiahao He, and others; colleagues in Dr. Patrick Schnable's lab: Yawei Li, Dr. Yan Zhou; colleagues in Dr. Yanhai Yin's lab: Dr. Trevor Nolan, Dr. Ping Wang, Dr. Hongqing Guo, Dr. Mitch Elmore, etc.; and colleagues in Dr. Jianming Yu's lab: Jialu Wei, Gregory Schoenbaum, and the undergraduate students who helped with data collection in the field. I also want to thank Dr. Maria G. Salas Fernandez and Juan S Panelo from the Department of Agronomy for their support in sorghum-related research. I thank my friends Dr. Kai Liu, Dr. Kailao Wang, Dr. Qi Mu, Dr. Si Chen, Yunjiao Zhu, Dr. Xiaoqian Mu, and others for making my time at Iowa State University a wonderful experience.

Most importantly, I thank my parents and sister for their love, understanding, patience, and support through the years; my boyfriend Dongyi Wang for his caring and encouragement. I dedicate my dissertation to my dear parents Mr. Zhangming Xiang and Mrs. Zhangchai Liu, who have provided continuous support in every step I took in pursuing my dream.

ABSTRACT

The ability to correlate morphological traits of plants with their genotypes plays an important role in plant phenomics research. However, traditional plant phenotyping is time-consuming, labor-intensive, and prone to human errors. This dissertation documents my innovative research in high-throughput robotic plant phenotyping for sorghum and maize plants using 3D machine vision and convolutional neural networks.

Sorghum is an important grain crop and a promising feedstock for biofuel production due to its excellent drought tolerance and water use efficiency. The 3D surface model of a plant can potentially provide an efficient and accurate way to digitize plant architecture and accelerate sorghum plant breeding programs. A non-destructive 3D scanning system using a commodity depth camera was developed to take side-view images of plants at multiple growth stages. A 3D skeletonization algorithm was developed to analyze the plant architecture and segment individual leaves. Multiple phenotypic parameters were obtained from the skeleton and the reconstructed point cloud including plant height, stem diameter, leaf angle, and leaf surface area. These image-derived features were highly correlated with the ground truth. Additionally, the results showed that stem volume was a promising predictor of shoot fresh weight and shoot dry weight.

To address the challenges of in-field imaging for plant phenotyping caused by variable outdoor lighting, wind conditions, and occlusions of plants. A customized stereo module, namely PhenoStereo, was developed for acquiring high-quality image data under field conditions. 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 Region Convolutional Neural Network for detecting stalk contours and a Semi-Global Block Matching stereo matching algorithm for generating disparity maps. The system-derived stem diameters were highly correlated with the ground truth.

Additionally, PhenoStereo was used to quantify the leaf angle of maize plants under field conditions. Multiple tiers of PhenoStereo camera were mounted on PhenoBot 3.0, a robotic vehicle designed to traverse between pairs of agronomically spaced rows of crops, to capture side-view images of maize plants in the field. An automated image processing pipeline (AngleNet) was developed to detect each leaf angle as a triplet of keypoints in two-dimensional images and extract quantitative data from reconstructed 3D models. AngleNet-derived leaf angles and their associated internode heights were highly correlated with manually collected ground-truth measurements.

The dissertation investigates and develops automated computer-vision-based robotic systems for plant phenotyping under controlled environments and in field conditions. In particular, a stereo module was customized and utilized to acquire high-quality image data for in-field plant phenotyping. With high-fidelity reconstructed 3D models and robust image processing algorithms, a series of plant-level and organ-level phenotypic traits of sorghum and maize plants were accurately extracted. The 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. The proposed approaches in this dissertation provide efficient alternatives to traditional phenotyping that could potentially accelerate breeding programs for improved plant architecture.

CHAPTER 1. GENERAL INTRODUCTION

1.1. Background

Improving the efficiency of plant breeding and crop production is crucial to meet the increasing food and energy demands of over nine billion global population by 2050 (Tilman et al., 2011). 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 and stress-tolerant crop species (Li et al., 2014). The robust crop species could be developed more rapidly than what is currently possible by analyzing the relationship between genotypes and phenotypes under various growing environments (Furbank and Tester, 2011). Boosted by technological developments such as high-throughput DNA sequencing, the efficiency of genotyping has been greatly improved (Phillips, 2010). However, the capability of phenotyping is still a bottleneck for dissecting the genetics of quantitative traits such as plant height, leaf angle, leaf area, and stem diameter (White et al., 2012). The conventional phenotyping practice tends to be time-consuming, labor-intensive, and prone to human errors, whereas most phenotypes were obtained in destructive ways or through manual observations. Thus, it is vitally important to develop techniques for collecting crop phenotypic data with higher accuracy and efficiency.

1.2. Proximal Sensing for High-Throughput Plant Phenotyping

The process that transforms raw sensor data into meaningful phenotypic data is essentially machine perception. Plant traits that can be characterized using high-throughput plant phenotyping (HTPP) technologies are largely limited by the available proximal sensing technologies. Plant traits can be quantified with proximal sensing at different levels of system integration (i.e., plant-level and organ-level). Plant-level traits are often generic and can be applied to a wide range of crop species and crop growth stages. Such traits include plant morphological measurements (e.g., height, width, volume, and surface area of the canopy) and plant physiological indicators (e.g., vegetation indices and canopy temperature), which require low-level machine perception. Organ-level traits are used to analyze individual organs of plants, including stem diameter, leaf angle, area of individual leaves, etc. However, due to the small sizes and occlusion issues in crop fields, machine perception at organ level remains as a challenging research area.

The majority of proximal sensors used in HTPP are imaging-based (Li et al., 2014). Some key imaging techniques for plant phenotyping include visible imaging, fluorescence imaging, thermal imaging, spectroscopic imaging, and LiDAR. In addition to imaging sensors, other proximal sensors often provide an average response in the field of view. Such sensors include ultrasonic and laser distance sensors, NDVI spectrometers (e.g., Crop Circle), and infrared radiometers. This section provides a thorough review of widely used proximal sensing techniques and their applications in HTPP.

1.2.1. RGB Vision

Digital red-green-blue (RGB) cameras are the most accessible sensors for plant phenotyping. Charged-coupled device (CCD) and complementary metal-oxide-semiconductor (CMOS) are the two major types of image sensors sensitive in the visible and near infrared electromagnetic spectral range. The three color channels are typically realized by placing a Bayer filter mosaic in alignment with the photosensor array. RGB cameras have been widely used on most HTPP platforms (Kicherer et al., 2017; Virlet et al., 2017). The RGB images with intensity values in red, green, and blue color channels provide the spatial information and color features of a scene. Applications based on RGB vision including disease detection and classification (DeChant et al., 2017), canopy cover estimation (Liu & Pattey, 2010), and weed detection (Abdalla, Cen, El-manawy, et al., 2019; Abdalla, Cen, Wan, et al., 2019). Suitable for outdoors applications, RGB cameras have the advantages of high spatial resolutions, fast acquisition rates, and low costs. However, the use of RGB imaging in the field is limited by complex background, wind, occlusion, and lighting variation. The disparities in image quality have affected the performance of the imaging analysis and pose challenges for algorithm development. Recent advancements in deep convolutional neural networks (CNNs) have provided state-of-the-art solutions to overcome these challenges in RGB image analysis. For example, DeepSeedling was developed to count cotton plant seedlings in the field (Jiang et al., 2019). The system-derived seedling counts were highly correlated ($r = 0.99$) with the ground truth. Besides plant counting, the applications of CNNs were further extended to count plant organs such as wheat spikes, corn tassels (Lu et al., 2017), rice panicles (Xiong et al., 2017), and cotton bolls (Li et al., 2017). Several studies have developed CNN-based approaches for detection and analysis of wheat spikes (Hasan et al., 2018; Sadeghi-Tehran et al., 2019; Xiong et al., 2019). Despite the challenging field conditions, the proposed deep learning models achieved high detection accuracy and satisfactory performance. CNN-based approaches demand adequate annotated datasets to ensure development and evolution. However, labeling diverse datasets is a labor-intensive task and becomes the major limiting factor for using deep learning in plant phenotyping. A recent review provides a comprehensive overview of the applications of CNNs in plant phenotyping applications (Jiang & Li, 2020).

1.2.2. Stereo Vision

Stereo vision, which reconstructs a 3D model from multi-view images, offers an inexpensive, efficient, and reasonably accurate solution for infield phenotyping of 3D plant morphology. Multi-view stereo systems combine multiple-view stereo (MVS) and structure-from-motion (SfM) to reconstruct 3D models of plants from images. Nguyen et al. (2016) developed an infield phenotyping system that utilized 16 color stereo vision cameras for capturing multi-view images of eggplants. An arc-shaped superstructure was designed to deploy the cameras and texture-enhancing structured illumination modules. Plant height, leaf number, leaf area, and plant biomass can be estimated by the system. The MVS-SfM method can generate dense point clouds of plants, but the major limitation is the intensive computation required for dense 3D reconstruction. Binocular stereo vision is being considered an alternative option to overcome limitations of the 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) developed a vertically moving boom system to scan infield energy sorghum plants with stereo cameras. Phenotypic traits such as leaf length, leaf width, and stem diameter were computed. Bao et al. (2019) developed a stereo-vision-based robotic system for tall dense canopy crops in the field. Stereo cameras were positioned at multiple heights on an extension rig to perform side-viewing imaging. Several morphological traits were quantified and found to be accurate and highly repeatable. Stereo vision has the advantages of high resolutions and low costs, however, the performance is affected by textureless regions, sunlight variation, and wind (Bao et al., 2019; Dandrifosse et al., 2020).

1.2.3. Time-of-Flight Cameras

A Time-of-Flight (ToF) camera is another widely used depth imaging sensor for infield plant phenotyping. Distances between an object and an image sensor are measured by calculating the time or phase shift of a light pulse that has travelled to and from the object. ToF cameras were successfully used for plant architecture characterization (Bao et al., 2019; Jiang et al., 2016), biomass estimation (Mortensen et al., 2018), fruit detection (Gené-Mola et al., 2019), and crop recognition and monitoring (Andújar et al., 2016). Low costs and high frame rates are the advantages of ToF cameras, whereas the major weakness is its poor performance under strong sunlight. To alleviate this issue, a sunshade was adopted to reduce the sunlight intensity in outdoor applications (Gai et al., 2020; Mortensen et al., 2018). Recent advancements in commercial-grade depth sensors provide an inexpensive solution for measuring plant architecture (Hu et al., 2018). For example, the Microsoft Kinect v2 camera, with upgraded color and depth resolutions, has shown promising performance in plant phenotyping. Bao et al. (2019) investigated the utility of side-viewing Kinect v2 cameras for characterization of corn plant architecture under field conditions at approximately sunset. A skeleton-based algorithm was developed to analyze the plant architecture and extract shoot architecture phenotypes. This work showed that the system was robust and accurate in acquiring several important traits including plant height, plant orientation, and leaf angle. Andújar et al. (2016) described a method for 3D modeling of cauliflower plants with a Kinect-based automated system. This method relied on the creation of a 3D point cloud in real time and the subsequent generation of a solid surface model. Strong correlations between measurements from the models and actual structural parameters were found.

1.2.4. Range Sensing

Light detection and ranging (LiDAR) is one of the most representative remote-sensing techniques for field-based phenotyping. LiDAR measures distances by illuminating an object and analyzing the reflected light source. In recent years, LiDAR has been widely integrated with field-based phenotyping systems and used for canopy reconstruction (Garrido et al., 2015; Jimenez-Berni et al., 2018), plant area density estimation (Hosoi & Omasa, 2009), plant height measurements (S. Sun et al., 2017; X. Wang et al., 2018), biomass estimation (Jimenez-Berni et al., 2018), nitrogen status assessment (Eitel et al., 2014), and plant growth analysis (Sun et al., 2018). For example, Qiu et al. (2019) reconstructed 3D models of corn plants by mounting a 3D LiDAR on a mobile ground robot. In their study, the robot with the 3D LiDAR moved in alleyways to collect data of corn plant rows. The point clouds were registered and merged by using the Iterative Closest Points algorithm with the assistance of landmarks in the field. Morphological traits including row spacing and plant height were derived from the 3D models. Besides canopy-level morphological traits, high-resolution 3D

LiDAR sensors have also provided the potential for organ-level traits, such as cotton bolls (Sun et al., 2018), sorghum panicles (Malambo et al., 2019), and corn stems (Jin et al., 2020). A terrestrial LiDAR sensor mounted on a sprayer has allowed the automated detection and morphological measurement of sorghum panicles using density-based clustering of LiDAR data (Lonesome Malambo & Heatwole, 2020). Cotton plant internode detection is feasible using terrestrial LiDAR data collected under field conditions (Sun et al., 2019). A Laplacian-based skeletonization algorithm and minimum-spanning-tree-based method were developed to analyze the structure of cotton plants. Thus, LiDAR sensors can be integrated into different plant phenotyping systems to characterize a wide range of phenotypes.

LiDAR provides detailed 3D distributions of plant canopies by structured light projection, thus providing high-resolution surface models and high accurate estimates of structural parameters. Some shortcomings of LiDAR sensors for plant phenotyping include: 1) low temporal resolution (Dandrifosse et al., 2020), i.e., takes long scanning time to obtain dense point cloud of plant canopy; 2) lack of color information in the measurement, but this can be solved by integrating other kinds of imaging sensors such as RGB cameras; and 3) for field-based plant phenotyping, LiDAR is sensitive to the noise caused by wind, insects, rain, and small panicles in the air (Perez-sanz et al., 2017). Though with these shortcomings, some new technologies such as high-density LiDAR, Full-waveform LiDAR, and Hyperspectral LiDAR have shown a great potential for the next-generation LiDAR-based plant phenotyping (Lin, 2015).

An ultrasonic sensor is a range detector, which has long been utilized for the evaluation of plant height (Sharma et al., 2016; Sui et al., 2012). The working principle of ultrasonic sensors is similar to that of LiDARs, whereas the difference is that an ultrasonic sensor emits sound waves to measure distance. Ultrasonic sensors have the advantages of low costs, high sampling rates, and simple communication interface. On the other hand, the sensors can suffer from materials and surfaces that can absorb sound waves, and they are only suitable for short-range sensing. Yuan et al. (2018) developed a multi-sensor phenotyping platform equipped with ultrasonic sensors and LiDARs. The study of measuring wheat plant height using the system showed that LiDAR provided better results than ultrasonic sensors. At a higher cost, the measurement produced by LiDAR is more accurate because of the higher spatial resolution (Qiu et al., 2018).

1.2.5. Spectral Sensing

Some other imaging techniques have been used in numerous plant phenotyping platforms, including spectral sensors and thermal imaging. Spectral sensors assess the spectral response from objects in a number of spectral bands including visible (400-700 nm), near infrared (700-1200 nm), and short-wavelength infrared regions (1400-3000 nm). Some commercial spectral sensors such as GreenSeeker (Trimble Inc., CA, USA) and Crop Circle (Holland Scientific Inc., NE, USA) were widely used by researchers to form vegetative indices such as normalized difference vegetation index (NDVI) (Barker et al., 2016; Bayati & Fotouhi, 2018; Crain et al., 2016). The high-resolution spectral measurements make it a promising tool for the detection of disease and abiotic stresses (Behmann et al., 2015; Corti et al., 2017; Mahlein et al., 2017; J. Sun et al., 2017; Thomas et al., 2018; Wang et al., 2019). Spectral image analysis is usually performed using appropriate statistical/machine learning algorithms to select significant features and build robust prediction models. These algorithms include least squares regression, support vector machine, principal components analysis, random forest, and neural networks. Furthermore, ensemble learning can be implemented to aggregate the strengths of multiple machine learning approaches (Moghimi et al., 2018). Thermal imaging, which can detect emitted infrared radiation in the mid- and long-wave infrared range from 3 to 12 um, allows for measuring temperature distribution on plant canopies. These different sensors have been integrated into phenotyping platforms to assess various agronomic traits. For example, Field Scanalyzer, an automated gantry-based field phenotyping platform, carrying a dedicated sensor array consisting of hyperspectral, thermal infrared, RGB, and chlorophyll fluorescence cameras, as well as 3D laser scanners (Virlet et al., 2017). The sensors array enables detailed non-invasive measurements of wheat plants including NDVI, canopy temperature, wheat head detection, plant height, and chlorophyll fluorescence. BreedVision is another multi-sensor platform developed for field-based plant phenotyping. Various optical sensors like hyperspectral imaging, 3D imagers and color cameras were used to collect spectral and morphological information of plants; and phenotypic traits including plant moisture content, tiller density, and lodging were measured from the multi-sensor fusion models (Busemeyer et al., 2013).

1.3. Research Objectives

Sorghum and maize plants are among the most important grain crops, and promising feedstocks for biofuel production due to their excellent drought tolerance (Xin et al., 2008) and water use efficiency (Borrell et al., 2014). If phenotypes such as plant height, stem diameter, leaf angle, and leaf area can be deduced from the 3D plant architecture without destroying the plant, then detailed evaluations of these parameters through time can help study the genetic basis of complex traits over a variety of genotypes, and hence improve the efficiency of breeding programs (Li et al., 2014). Moreover, non-destructive biomass estimation can be important to dissect the genetic architecture and variability of sorghum biomass yield under different stress conditions. However, in reality, phenotyping sorghum and maize plants via imaging remains a particularly challenging task due to occlusions caused by overlapping leaves and tillers, often leading to the requirement of user interaction to facilitate the image analysis (Pound et al., 2014; Salas Fernandez et al., 2017).

To automate a process that can accurately measure organ-level traits of sorghum and maize plants in the field has been further challenging. Typically, the organ-level traits such as stem diameter and leaf angle are measured manually in the field, which is a tedious and error-prone practice. Several studies have investigated the suitability of vision-based systems for field-based phenotypic traits characterization, but the methods usually require human intervention. Recent advances in machine learning offered new approaches for reliable object detection in some image-based deep learning tasks. These studies identified and detected the plant organs in 2D images and performed well. However, the performance could be further improved by acquiring higher-quality images.

The overall goal of this research is to develop methodologies to characterize architectural traits of sorghum and maize plants using side-view 3D imaging. The research project reposted in this dissertation has two phases. The first phase was the development of a non-destructive 3D scanning system using a commodity depth camera. The reviewed previous studies have mainly focused on relatively simple cases where sorghum plants have no overlapped leaves or tillers. The overall goal of this study is to automatically extract sorghum plant architectural parameters via 3D point cloud analysis. The specific objectives are to: 1) reconstruct the 3D surface model of the imaged plants; 2) develop a robust data processing pipeline to characterize the plant architecture; 3) extract phenotypic data including plant height, stem diameter, leaf angle, and leaf area; and 4) explore the possibilities of using the extracted traits for plant biomass estimation.

The second phase was the study of the feasibility of stereo imaging for measuring plant architectural traits of sorghum and maize plants in the field. The characterization of two important plant architecture traits, including stem diameter and leaf angle, were studied and investigated. Measuring the stem diameter of sorghum plants has been identified as a challenging sensing task to automate in the field due to the complexity of the imaging object and the unstructured environment. The leaf angle characterization of field-grown maize plants is regarded as one of the most challenging phenotyping tasks because of the substantial overlap and occlusion in maize plant canopies; and moreover, variable stem and leaf orientations relative to the camera can affect their visibility near a leaf collar. To relieve these stresses, a portable stereo imaging module, namely PhenoStereo, was developed for acquiring high-quality stereoscopic images in the field. A series of image processing algorithms making use of 3D reconstruction and CNN-based object and keypoint detection techniques were used to automatically detect and characterize organ-level traits of sorghum and maize plants under field conditions. The specific objectives for phenotyping each crop species are listed below:

Research Objectives of Field-Based Stem Diameter Detection and Characterization of Sorghum Plants:

1) acquire high-quality images of sorghum plants in the field using our customized stereo module;

2) develop an automated image processing pipeline to detect the stems of sorghum plants and measure their stem diameter in reconstructed 3D models; and

3) evaluate the performance of stem instance segmentation and stem diameter estimation.

Research Objectives of Field-Based Robotic Leaf Angle Detection and Characterization of Maize Plants:

1) develop a customized stereo-vision-based plant phenotyping platform to image maize plant canopy at different heights in the field;

2) develop an automated image processing pipeline to detect the regions and keypoints of interest and characterize plant architecture via a 3D modeling process and derive two important traits: leaf angle and its associated node height;

3) evaluate the performance of the proposed approach by comparing system-derived measurements with ground truth; and

4) explore the effectiveness of the newly developed system for characterizing leaf angle variations in different maize inbred lines in the field.

1.4. Dissertation Outline

The dissertation comprises a compilation of three journal articles that document my studies in two strands of research in automated plant phenotyping: (1) morphological traits extraction for sorghum plants under controlled environments, and (2) field-based organ-level traits detection and characterization of sorghum and maize plants. Chapter 1 provides a general introduction followed by the research objectives. The first article (Chapter 2) presents the study of automated morphological traits extraction for sorghum plants via 3D point cloud data analysis. The second and third articles focus on characterizing organ-level phenotypic traits in the field. The second article (Chapter 3) describes a high-throughput stereo vision system for field-based plant phenotyping and investigates measuring stem diameter of sorghum plants in the field using the stereo vision system. The third article (Chapter 4) details field-based robotic leaf angle detection and characterization of maize plants using stereo vision and deep convolutional neural networks. General conclusions are drawn in Chapter 5 along with a list of recommendations for future research.

References

Abdalla, A., Cen, H., El-manawy, A., & He, Y. (2019). Infield oilseed rape images segmentation via improved unsupervised learning models combined with supreme color features. *Computers and Electronics in Agriculture*, *162*, 1057–1068. https://doi.org/10.1016/j.compag.2019.05.051

Abdalla, A., Cen, H., Wan, L., Rashid, R., Weng, H., Zhou, W., & He, Y. (2019). Fine-tuning convolutional neural network with transfer learning for semantic segmentation of ground-level oilseed rape images in a field with high weed pressure. *Computers and Electronics in Agriculture*, *167*, 105091. https://doi.org/10.1016/j.compag.2019.105091

Andújar, D., Dorado, J., Fernández-Quintanilla, C., & Ribeiro, A. (2016). An Approach to the Use of Depth Cameras for Weed Volume Estimation. *Sensors*, *16*(7), 972. https://doi.org/10.3390/s16070972

Andújar, D., Ribeiro, A., Fernández-Quintanilla, C., & Dorado, J. (2016). Using depth cameras to extract structural parameters to assess the growth state and yield of cauliflower crops. *Computers and Electronics in Agriculture, 122*, 67–73. https://doi.org/10.1016/j.compag.2016.01.018

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.

Bao, Y., Tang, L., Srinivasan, S., & Schnable, P. S. (2019). Field-based architectural traits characterisation of maize plant using time-of-flight 3D imaging. *Biosystems Engineering*, *178*, 86–101.

Barker, J., Zhang, N., Sharon, J., Steeves, R., Wang, X., Wei, Y., & Poland, J. (2016). Development of a field-based high-throughput mobile phenotyping platform. *Computers and Electronics in Agriculture*, *122*, 74–85. https://doi.org/10.1016/j.compag.2016.01.017

Bayati, M., & Fotouhi, R. (2018). A Mobile Robotic Platform for Crop Monitoring. *Advances in Robotics & Automation*, *07*(01). https://doi.org/10.4172/2168-9695.1000186

Behmann, J., Mahlein, A. K., Rumpf, T., Römer, C., & Plümer, L. (2015). A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. In *Precision Agriculture* (Vol. 16, Issue 3, pp. 239–260). Kluwer Academic Publishers. https://doi.org/10.1007/s11119-014-9372-7

Borrell, A.K., Mullet, J.E., George-Jaeggli, B., van Oosterom, E.J., Hammer, G.L., Klein, P.E., Jordan, D.R., 2014. Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. J. Exp. Bot. 65, 6251–6263. https://doi.org/10.1093/jxb/eru232

Busemeyer, L., Mentrup, D., Möller, K., Wunder, E., Alheit, K., Hahn, V., Maurer, H. P., Reif, J. C., Würschum, T., & Müller, J. (2013). BreedVision—A multi-sensor platform for non-destructive field-based phenotyping in plant breeding. *Sensors*, *13*(3), 2830–2847. https://doi.org/10.3390/s130302830

Corti, M., Marino Gallina, P., Cavalli, D., & Cabassi, G. (2017). Hyperspectral imaging of spinach canopy under combined water and nitrogen stress to estimate biomass, water, and nitrogen content. *Biosystems Engineering*, *158*, 38–50. https://doi.org/10.1016/j.biosystemseng.2017.03.006

Crain, J. L., Wei, Y., Barker, J., Thompson, S. M., Alderman, P. D., Reynolds, M., Zhang, N., & Poland, J. (2016). Development and deployment of a portable field phenotyping platform. *Crop Science*, *56*(3), 965–975. https://doi.org/10.2135/cropsci2015.05.0290

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*. https://doi.org/10.3389/fpls.2020.00096

DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E. L., Yosinski, J., Gore, M. A., Nelson, R. J., & Lipson, H. (2017). Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology*, *107*(11), 1426–1432. https://doi.org/10.1094/PHYTO-11-16-0417-R

Furbank, R.T., Tester, M., 2011. Phenomics – technologies to relieve the phenotyping bottleneck. Trends Plant Sci. 16, 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*, *37*(1), 35–52. https://doi.org/10.1002/rob.21897

Garrido, M., Paraforos, D. S., Reiser, D., Arellano, M. V., Griepentrog, H. W., & Valero, C. (2015). 3D maize plant reconstruction based on georeferenced overlapping lidar point clouds. *Remote Sensing*, *7*(12), 17077–17096. https://doi.org/10.3390/rs71215870

Gené-Mola, J., Vilaplana, V., Rosell-Polo, J. R., Morros, J.-R., Ruiz-Hidalgo, J., & Gregorio, E. (2019). Multi-modal deep learning for Fuji apple detection using RGB-D cameras and their radiometric capabilities. *Computers and Electronics in Agriculture*, *162*, 689–698. https://doi.org/10.1016/j.compag.2019.05.016

Hasan, M. M., Chopin, J. P., Laga, H., & Miklavcic, S. J. (2018). Detection and analysis of wheat spikes using Convolutional Neural Networks. *Plant Methods*, *14*(1), 100. https://doi.org/10.1186/s13007-018-0366-8

Hosoi, F., & Omasa, K. (2009). Estimating vertical plant area density profile and growth parameters of a wheat canopy at different growth stages using three-dimensional portable lidar imaging. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(2), 151–158. https://doi.org/10.1016/j.isprsjprs.2008.09.003

Hu, Y., Wang, L., Xiang, L., Wu, Q., & Jiang, H. (2018). Automatic non-destructive growth measurement of leafy vegetables based on kinect. *Sensors*, *18*(3), 806. https://doi.org/10.3390/s18030806

Jiang, Y., & Li, C. (2020). Convolutional Neural Networks for Image-Based High-Throughput Plant Phenotyping: A Review. *Plant Phenomics*, *2020*, 1–22. https://doi.org/10.34133/2020/4152816

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

Jiang, Y., Li, C., Paterson, A. H., & Robertson, J. S. (2019). DeepSeedling: deep convolutional network and Kalman filter for plant seedling detection and counting in the field. *Plant Methods*, *15*(1), 141. https://doi.org/10.1186/s13007-019-0528-3

Jimenez-Berni, J. A., Deery, D. M., Rozas-Larraondo, P., Condon, A. T. G., Rebetzke, G. J., James, R. A., Bovill, W. D., Furbank, R. T., & Sirault, X. R. R. (2018). High throughput determination of plant height, ground cover, and above-ground biomass in wheat with LiDAR. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.00237

Jin, S., Su, Y., Gao, S., Wu, F., Ma, Q., Xu, K., Hu, T., Liu, J., Pang, S., Guan, H., Zhang, J., & Guo, Q. (2020). Separating the Structural Components of Maize for Field Phenotyping Using Terrestrial LiDAR Data and Deep Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(4), 2644–2658. https://doi.org/10.1109/TGRS.2019.2953092

Kicherer, A., Herzog, K., Bendel, N., Klück, H. C., Backhaus, A., Wieland, M., Rose, J. C., Klingbeil, L., Läbe, T., Hohl, C., Petry, W., Kuhlmann, H., Seiffert, U., & Töpfer, R. (2017). Phenoliner: A new field phenotyping platform for grapevine research. *Sensors (Switzerland), 17*(7). https://doi.org/10.3390/s17071625

Li, L., Zhang, Q., Huang, D., Li, L., Zhang, Q., Huang, D., 2014. A Review of Imaging Techniques for Plant Phenotyping. Sensors 14, 20078–20111. https://doi.org/10.3390/s141120078

Li, Y., Cao, Z., Xiao, Y., & Cremers, A. B. (2017). DeepCotton: in-field cotton segmentation using deep fully convolutional network. *Journal of Electronic Imaging*, *26*(5), 53028.

Lin, Y. (2015). LiDAR: An important tool for next-generation phenotyping technology of high potential for plant phenomics? In *Computers and Electronics in Agriculture* (Vol. 119, pp. 61–73). Elsevier B.V. https://doi.org/10.1016/j.compag.2015.10.011

Liu, Jiangui, & Pattey, E. (2010). Retrieval of leaf area index from top-of-canopy digital photography over agricultural crops. *Agricultural and Forest Meteorology*, *150*(11), 1485–1490. https://doi.org/10.1016/j.agrformet.2010.08.002

Lu, H., Cao, Z., Xiao, Y., Zhuang, B., & Shen, C. (2017). TasselNet: counting maize tassels in the wild via local counts regression network. *Plant Methods, 13*(1), 79.

Mahlein, A.-K., Kuska, M. T., Thomas, S., Bohnenkamp, D., Alisaac, E., Behmann, J., Wahabzada, M., & Kersting, K. (2017). Plant disease detection by hyperspectral imaging: from the lab to the field. *Advances in Animal Biosciences*, *8*(2), 238–243. https://doi.org/10.1017/s2040470017001248

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

Malambo, Lonesome, & Heatwole, C. D. (2020). Automated training sample definition for seasonal burned area mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *160*, 107–123. https://doi.org/10.1016/j.isprsjprs.2019.11.026

Moghimi, A., Yang, C., & Marchetto, P. M. (2018). Ensemble Feature Selection for Plant Phenotyping: A Journey from Hyperspectral to Multispectral Imaging. *IEEE Access, 6*, 56870–56884. https://doi.org/10.1109/ACCESS.2018.2872801

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

Nguyen, T. T., Slaughter, D. C., Maloof, J. N., & Sinha, N. (2016). Plant phenotyping using multi-view stereo vision with structured lights. *Proc. SPIE 9866, Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping*, *9866*, 986608.

Perez-sanz, F., Navarro, P. J., & Egea-cortines, M. (2017). Plant phenomics : an overview of image acquisition technologies and image data analysis algorithms. *GigaScience, February*, 1–18. https://doi.org/10.1093/gigascience/gix092

Phillips, R.L., 2010. Mobilizing Science to Break Yield Barriers. Crop Sci. 50, S-99. https://doi.org/10.2135/cropsci2009.09.0525

Pound, M.P., French, A.P., Murchie, E.H., Pridmore, T.P., 2014. Automated recovery of three-dimensional models of plant shoots from multiple color images. Plant Physiol. 166, 1688–98. https://doi.org/10.1104/pp.114.248971

Qiu, Q., Sun, N., Bai, H., Wang, N., Fan, Z., Wang, Y., Meng, Z., Li, B., & Cong, Y. (2019). Field-based high-throughput phenotyping for maize plant using 3d LIDAR point cloud generated with a “phenomobile.” *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.00554

Qiu, R., Wei, S., Zhang, M., Li, H., Sun, H., Liu, G., & Li, M. (2018). Sensors for measuring plant phenotyping: A review. *International Journal of Agricultural and Biological Engineering*, *11*(2), 1–17. https://doi.org/10.25165/j.ijabe.20181102.2696

Sadeghi-Tehran, P., Virlet, N., Ampe, E. M., Reyns, P., & Hawkesford, M. J. (2019). DeepCount: In-Field Automatic Quantification of Wheat Spikes Using Simple Linear Iterative Clustering and Deep Convolutional Neural Networks. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.01176

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, 2008–2022. https://doi.org/10.1104/pp.17.00707

Sharma, L. K., Bu, H., Franzen, D. W., & Denton, A. (2016). Use of corn height measured with an acoustic sensor improves yield estimation with ground based active optical sensors. *Computers and Electronics in Agriculture*, *124*, 254–262.

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

Sui, R., Fisher, D. K., & N. Reddy, K. (2012). Cotton Yield Assessment Using Plant Height Mapping System. *Journal of Agricultural Science*, *5*(1), p23. https://doi.org/10.5539/jas.v5n1p23

Sun, J., Shi, S., Gong, W., Yang, J., Du, L., Song, S., Chen, B., & Zhang, Z. (2017). Evaluation of hyperspectral LiDAR for monitoring rice leaf nitrogen by comparison with multispectral LiDAR and passive spectrometer. *Scientific Reports*, *7*(1), 1–9. https://doi.org/10.1038/srep40362

Sun, S., Li, C., Paterson, A., Adhikari, J., Robertson, J., & Meng, C. (2019). Automated plant node detection using terrestrial LiDAR data under field conditions. *2019 Boston, Massachusetts July 7- July 10, 2019*, 1--. https://doi.org/10.13031/aim.201900598

Sun, S., Li, C., Paterson, A. H., Jiang, Y., Xu, R., Robertson, J. S., Snider, J. L., & Chee, P. W. (2018). In-field high throughput phenotyping and cotton plant growth analysis using LiDAR. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.00016

Thomas, S., Behmann, J., Steier, A., Kraska, T., Muller, O., Rascher, U., & Mahlein, A. K. (2018). Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform. *Plant Methods*, *14*(1), 45. https://doi.org/10.1186/s13007-018-0313-8

Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proc. Natl. Acad. Sci. U. S. A. 108, 20260–4. https://doi.org/10.1073/pnas.1116437108

Virlet, N., Sabermanesh, K., Sadeghi-Tehran, P., & Hawkesford, M. J. (2017). Field Scanalyzer: an automated robotic field phenotyping platform for detailed crop monitoring. *Functional Plant Biology*, *44*(1), 143–153. https://doi.org/10.1071/FP16163

Wang, D., Vinson, R., Holmes, M., Seibel, G., Bechar, A., Nof, S., & Tao, Y. (2019). Early detection of tomato spotted wilt virus by hyperspectral imaging and outlier removal auxiliary classifier generative adversarial nets (OR-AC-GAN). *Scientific reports*, *9*(1), 1-14. https://doi.org/10.1038/s41598-019-40066-y

White, J.W., Andrade-Sanchez, P., Gore, M.A., Bronson, K.F., Coffelt, T.A., Conley, M.M., Feldmann, K.A., French, A.N., Heun, J.T., Hunsaker, D.J., Jenks, M.A., Kimball, B.A., Roth, R.L., Strand, R.J., Thorp, K.R., Wall, G.W., Wang, G., 2012. Field-based phenomics for plant genetics research. F. Crop. Res. 133, 101–112. https://doi.org/10.1016/J.FCR.2012.04.003

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, 103. https://doi.org/10.1186/1471-2229-8-103

Xiong, H., Cao, Z., Lu, H., Madec, S., Liu, L., & Shen, C. (2019). TasselNetv2: In-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods*, *15*(1), 150. https://doi.org/10.1186/s13007-019-0537-2

Xiong, X., Duan, L., Liu, L., Tu, H., Yang, P., Wu, D., Chen, G., Xiong, L., Yang, W., & Liu, Q. (2017). Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods*, *13*(1), 104. https://doi.org/10.1186/s13007-017-0254-7

Yuan, W., Li, J., Bhatta, M., Shi, Y., Baenziger, P., & Ge, Y. (2018). Wheat Height Estimation Using LiDAR in Comparison to Ultrasonic Sensor and UAS. *Sensors*, *18*(11), 3731. https://doi.org/10.3390/s18113731

CHAPTER 2. AUTOMATED MORPHOLOGICAL TRAITS EXTRACTION FOR SORGHUM PLANTS VIA 3D POINT CLOUD DATA ANALYSIS

Modified from a manuscript published in *Computers and Electronics in Agriculture*

Lirong Xiang[1], Yin Bao[1], Lie Tang[1, *], Diego Ortiz[2], Maria G. Salas-Fernandez[2]

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

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

Abstract

The ability to correlate morphological traits of plants with their genotypes plays an important role in plant phenomics research. However, measuring phenotypes manually is time-consuming, labor intensive, and prone to human errors. The 3D surface model of a plant can potentially provide an efficient and accurate way to digitize plant architecture. This study focused on the extraction of morphological traits at multiple developmental timepoints from sorghum plants grown under controlled conditions. A non-destructive 3D scanning system using a commodity depth camera was implemented to capture sequential images of a plant at different heights. To overcome the challenges of overlapping tillers, an algorithm was developed to first search for the stem in the merged point cloud data, and then the associated leaves. A 3D skeletonization algorithm was created by slicing the point cloud along the vertical direction, and then linking the connected Euclidean clusters between adjacent layers. Based on the structural clues of the sorghum plant, heuristic rules were implemented to separate overlapping tillers. Finally, each individual leaf was automatically segmented, and multiple parameters were obtained from the skeleton and the reconstructed point cloud including plant height, stem diameter, leaf angle, and leaf surface area. The results showed high correlations between the manual measurements and the estimated values generated by the system. Statistical analyses between biomass and extracted traits revealed that stem volume was a promising predictor of shoot fresh weight and shoot dry weight, and the total leaf area was strongly correlated to shoot biomass at early stages.

Keywords. 3D point cloud data, phenotyping, plant morphology, skeletonization, sorghum

2.1. Introduction

Improving the efficiency of plant breeding and crop production is crucial to the success of meeting the increasing food and energy demands of over nine billion global population by 2050 (Tilman et al., 2011). High yielding, stress-tolerant plants could be developed more rapidly than is currently possible by analyzing the relationship between genotypes and phenotypes under various growth environments (Furbank & Tester, 2011). Boosted by technological developments such as high-throughput DNA sequencing, the efficiency of genotyping has been greatly improved (Phillips, 2010). However, the capability of phenotyping is still a bottleneck for dissecting the genetics of quantitative traits such as plant height, leaf angle, leaf area, and biomass (White et al., 2012). The conventional phenotyping practice tends to be time-consuming, labor-intensive, and prone to human errors. Thus, it is vitally important to develop techniques for collecting crop phenotypic data with higher accuracy and efficiency.

Though two-dimensional (2D) images have been widely used in image-based phenotyping techniques, 2D methods struggle to accurately reflect three-dimensional (3D) quantities (Gibbs et al., 2018; Kaminuma et al., 2004). Previous works has demonstrated that analyzing plants in 3D space offers greater robustness and accuracy (An et al., 2017; Apelt et al., 2015). In addition to the quantitative description, 3D imaging enables measurements of more traits than what is possible with 2D images such as the volume of fruits (Paulus, Behmann, et al., 2014). To date, advanced computer vision techniques have employed 3D information to quantify plant traits (L. Li et al., 2014). Generally, there are two approaches for 3D image-based phenotyping. In one approach, 3D surface models are reconstructed by 2D images taken from multiple views, which is referred to as the passive methods. Stereo vision is one of the passive imaging techniques that reconstruct 3D canopies by synthesizing multi-view images. This method has been successfully used for canopy model reconstruction (Bao et al., 2019; Salas Fernandez et al., 2017) and growth estimation (Rovira-Más et al., 2005). However, the performance is affected by the lack of texture in the objects and sunlight variations (Frasson & Krajewski, 2010). Structure-from-Motion (SfM) is another commonly used technique for 3D reconstruction due to the relatively high accuracy of the sensors (Bellasio et al., 2012; Nguyen et al., 2015), but its major limitation is the intensive computation required for reconstruction (Jay et al., 2015). Another approach to 3D-based plant phenotyping is an active method that directly acquires 3D distribution of plant canopies by using external light sources (Jiang et al., 2018). Several 3D sensors have been used for 3D modeling of plants and leaves, such as Light Detection and Ranging (LiDAR), Time-of-Flight (ToF) cameras, and 3D laser scanners. LiDAR applications include canopy reconstruction (Garrido et al., 2015), and the estimation of LAI (Tang et al., 2014), plant height (S. Sun et al., 2017), volume and biomass (Eitel et al., 2014; Rosell Polo et al., 2009)). ToF cameras have been used for crop recognition (J. Li & Tang, 2018), crop monitoring (Andújar, Dorado, et al., 2016; Jiang et al., 2016), and plant traits measurement (Alenya et al., 2011; Chaivivatrakul et al., 2014). Finally, laser scanners have been successfully deployed for LAI mapping (Gebbers et al., 2011), plant organ classification (Paulus et al., 2013), and growth tracking of plants over time (Paulus, Dupuis, et al., 2014). High accuracy, high signal update rate, and strong robustness are the advantages of these devices, though, the major limitations are the high signal noise level and sensitiveness to environmental variations (e.g., sunlight, wind).

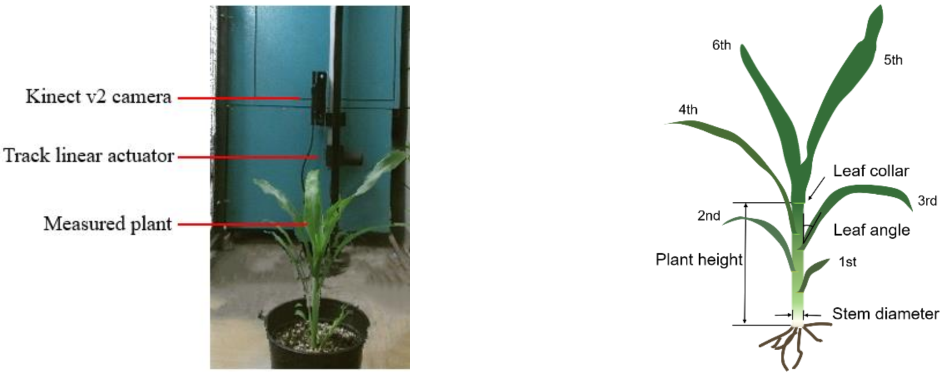
Recent advances in commercial-grade depth sensors provide a new solution for measuring plant architecture (Hu et al., 2018). The Microsoft Kinect v2 camera is one of the most commonly used depth sensors that, with upgraded color and depth resolutions, has performed well in plant phenotyping. McCormick et al. (2016) developed a 3D reconstruction system for sorghum (*Sorghum bicolor*) plants using a Kinect v2 camera with which plant architecture could be generated from a semi-automated data processing pipeline and shoot architecture phenotypes could be automatically measured. This work showed that measurements obtained with the Kinect depth camera are reliable for applications such as quantitative trait loci (QTLs) mapping. Andujar et al. (Andújar, Ribeiro, et al., 2016) described a method for 3D modeling of cauliflower plants with a Kinect-based automated system. This method relied on the creation of a 3D point cloud in real time, and the subsequent generation of a solid surface model. In their work, significant correlations between measurements from the models and actual structural parameters were found.

Sorghum is the fifth most important grain crop, and a promising feedstock for biofuel production due to its excellent drought tolerance (Xin et al., 2008) and water use efficiency (Borrell et al., 2014). If phenotypes such as plant height, stem diameter, and leaf area can be deduced from the 3D plant architecture without destroying the sorghum plant, then detailed evaluations of these parameters through time can help study the genetic basis of complex traits over a variety of genotypes, and hence improve the efficiency of breeding programs (L. Li et al., 2014). Moreover, non-destructive biomass estimation can be important to dissect the genetic architecture and variability of sorghum biomass yield under different stress conditions. However, in reality, phenotyping sorghum plants via imaging remains a particularly challenging task due to the occlusions caused by overlapping leaves and tillers, often leading to the requirement of user interaction to facilitate the image analysis (Pound et al., 2014; Salas Fernandez et al., 2017). The overall goal of this study is to automatically extract sorghum plant architectural parameters via 3D point cloud analysis. The specific objectives are to: (1) reconstruct the 3D surface model of the imaged plants, (2) develop a robust data processing pipeline to characterize the plant architecture, (3) extract phenotypic data including plant height, stem diameter, leaf angle, and leaf area, and (4) explore the possibilities of using the extracted traits for plant biomass estimation.

2.2. Materials and Methods

2.2.1. Experimental Design and Data Acquisition

A non-destructive imaging system was developed for data acquisition. It consisted of a Kinect v2 camera, a vertical track linear actuator, and a computer station (Figure 2.1). The Kinect v2 has a depth camera with a resolution of 512×424 pixels, an RGB camera of 1920×1080 pixels, and an infrared emitter. Kinect v2's depth measurement is based on Time-of-Flight (ToF) of light principle, where the distance between the emitter to the target is calculated from the travel time of the modulated light (Andújar, Dorado, et al., 2016). The Kinect camera was mounted on a vertically oriented and tracked linear actuator. The sorghum plant was positioned 0.9 meters away from the linear actuator, with the widest side of the canopy facing the camera. The system took RGB-D images at multiple heights while moving the camera at a speed of 40 mm/s. Starting from the same height, the system stored one image in every second with more than 90% of overlapped areas between consecutive images. A total of four images were saved for each plant based on the heights of the sorghum plants used in this project. Images were taken in the laboratory under typical office lighting conditions.



(a) (b)

Figure 2.1. System setup and manual measurements (a) The setup of the image acquisition station. (b) Illustration of the manual measurements of plant height, leaf angle, and stem diameter.

A total of 32 sorghum plants were grown in a growth chamber and studied to validate the system. These plants corresponded to four genotypes (PI 656029, 533839, 564163 and 655996) and eight biological replicates per genotype. The plants were sampled at multiple developmental timepoints: 21 (W3), 28 (W4), and 35 (W5) days after planting (DAP). Figure 2.2 shows the example images of one of the sampled sorghum plants at different growth stages. Plants were transported manually from the growth chamber to the imaging station. After taking images, plants were manually measured to evaluate the performance of the system. The ground truth data were collected by the same person. Plant height was measured from the soil to the collar of the last fully expanded leaf using a measuring tape. Stem diameter was measured using a caliper at the stem section right above the soil, and leaf angle was measured only for the leaves completely out of the whorl by a protractor (Figure 2.1b). Each week, eight plants (two per genotype) were randomly selected and cut at the base of the stem for leaf area and biomass measurements. Leaf area was measured by an LI-3000C Area Meter, and fresh biomass weight was obtained using an electronic balance. Subsequently, the plant was dried in an oven with a temperature of 65°C until the sample reached a constant weight. Total dry weight was then measured and recorded.

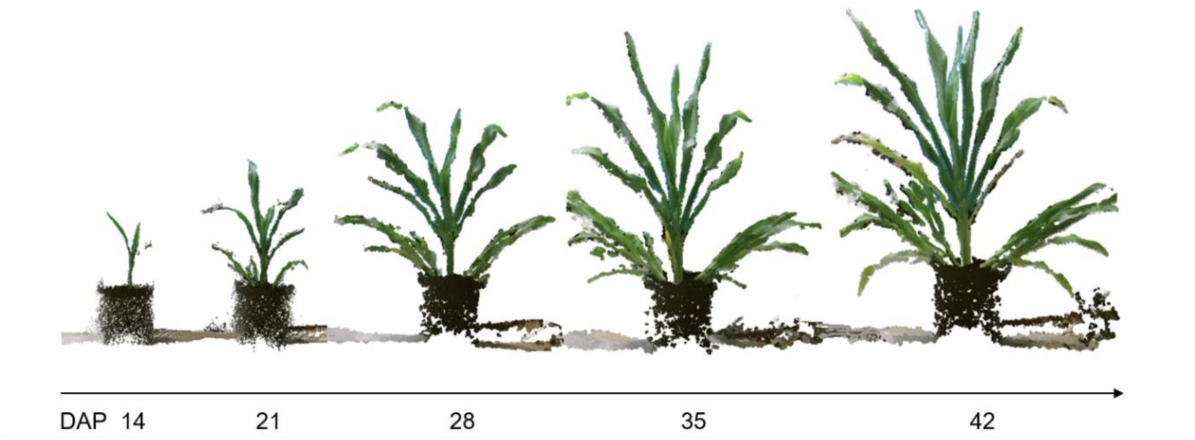


Figure 2.2. Point clouds of a sorghum plant at different growth stages, where DAP stands for Days After Planting.

2.2.2. Image Processing Pipeline

An automatic image processing pipeline was developed for determining plant architecture and characterizing phenotypic parameters. It comprised three sections: point cloud preprocessing, stem and leaf segmentation, and morphological traits extraction (Figure 2.3). Routines from Point Cloud Library (Rusu & Cousins, 2011) and OpenCV library (Bradski & Kaehler, 2008) were adopted to implement those steps.

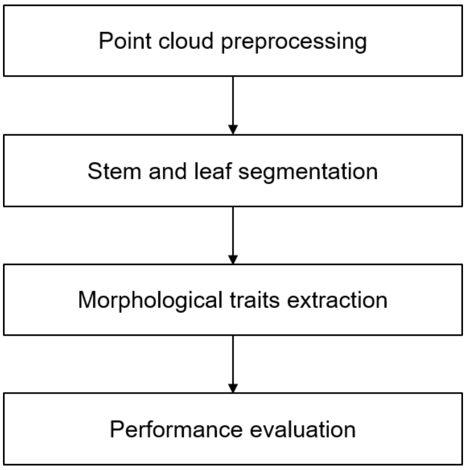


Figure 2.3. Main steps of the proposed automatic processing pipeline

2.2.2.1. Point Cloud Preprocessing

An overview of the point cloud preprocessing is shown in Figure 2.4. The four depth images taken at different heights were transformed into 3D point cloud using the equations listed as follows:

$$\mathrm{Z} = \mathrm{d}; \tag{2.1}$$

$$\mathrm{X} = (\mathrm{x} - c_x)/f_x * Z; \tag{2.2}$$

$$\mathrm{Y} = \left(\mathrm{y} - c_y\right)/f_y * Z; \tag{2.3}$$

where (x, y) is the 2D image point coordinates, d is the depth value of the 2D point. c_x and c_y represent the principal points along the camera X and Y axis. f_x and f_y are the focal lengths along the camera X, Y axis, respectively. (X, Y, Z) is the 3D coordinates of the corresponding point.

The point cloud data generated by the system contain some background objects (e.g., the supporting structure, the door, and walls of the room) which was removed by filtering out the 3D points outside a predefined bounding box containing the plant. The algorithm checks the X, Y, and Z coordinates of each point in the point cloud. Figure 2.5a shows the results after background removal, where the point clouds obtained from different heights are visualized with different colors.

Due to the limitation of the sensor, the raw data of the sorghum point cloud always contains a few sparsely distributed noise points which commonly have low point density. A radius-based outlier removal filter was implemented to clean the noise points. This filter takes into account the number of neighboring points (K) within a user-defined search radius r of all points, a point will be removed as noise if less than K neighbors are found (Dziubich et al., 2016). We selected r = 5 mm and K = 5 which effectively reduced the number of noise points without losing true surface points.

After background and noise removal, multiple neighboring point clouds were incrementally registered into a single point cloud (Figure 2.5b) by using the Iterative Closest Point (ICP) algorithm. The ICP algorithm aligns two point clouds by iteratively minimizing the Euclidean distances between corresponding points (Holz et al., 2015). The multi-view point cloud registration was done by finding the best transformation matrix between each consecutive clouds and converging these transformation matrices towards the first cloud's frame.

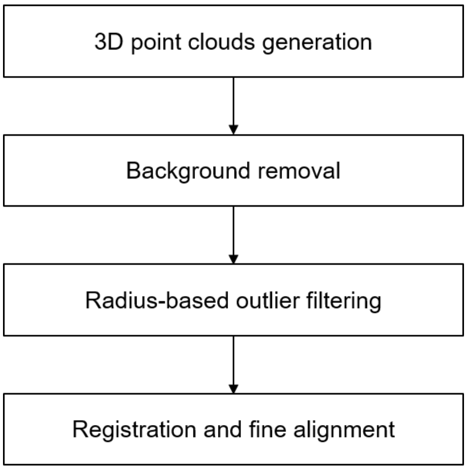


Figure 2.4. Flowchart of the algorithm for point cloud preprocessing

2.2.2.2. Stem and Leaf Segmentation

The phenotypic data extraction of the plant organs (i.e., leaves and stems) was based on segmentation. Previous work has developed and validated methods for plant components segmentation. One of them was to first remove the stem, and then separate every single leaf by clustering methods (Lu et al., 2017). However, this was based on a relatively simple plant architecture with leaves that were spatially separated from each other after filtering the stem. Another method used machine learning approaches to classify each point by point-level 3D features (McCormick et al., 2016; Sodhi et al., 2017), but the approach needed a manually labeled subset for training the classifier, which was time-consuming. The plants used in our experiment not only had leaves but also tillers close to the stem, and the leaves at the top of the canopy remained connected. To overcome these challenges, a skeletonization method was implemented to help build the structural clues of the whole plant. The algorithm was developed to first search for the stem by Hough line transform, and then segment each individual leaf.

Step 1: Soil Detection

The soil detection started with ground detection and the subsequent soil level estimation based on the ground level and the pot height. For ground detection, the Random Sample Consensus (RANSAC) algorithm (Fischler & Bolles, 1981) was used to fit a plane model. The method was applied to the full point cloud to find all the points that supported a plane model. Similarly, after ground detection, a soil detection process based on the RANSAC method was employed. We defined the direction perpendicular to the ground plane pointing upwards as the X-axis. The pot height was 0.20m and the soil was always inside the pot. The soil detection only considered those points with an X-value from X_{ground} + 0.15m to X_{ground} + 0.20m, thus increasing the ratio of soil to all the points and improving the detection accuracy. The inliers of the plane model were extracted as the soil and the soil level X_{soil} was estimated by averaging the X coordinates of all the inliers. The points above the soil were extracted as the region of interest and were used for the phenotypic traits extraction. Figure 2.5c shows the point cloud with the ground inliers in blue and soil inliers in red.

Step 2: Point Cloud Skeletonization

We introduced a point cloud skeletonization method to analyze the sorghum plant architecture. The 3D skeletonization algorithm was developed by slicing the point cloud along the X-axis (Figure 2.5d) and linking the connected Euclidean clusters (Rusu & Cousins, 2011) between adjacent layers. The generated skeleton was mapped to an undirected graph with each node denoting the cluster's centroid and the edge weight denoting the Euclidean distance between the two connected nodes. For each node, its neighbors and indices of the points in the cluster were stored. Due to inter-plant occlusion and boundary noise, the initial skeleton graph may include loops and redundant spurious branches, thus clean skeletons were extracted by removing loops using Kruskal’s algorithm (Kruskal, 1956) and pruning spurious branches. The process of skeleton pruning was based on thresholding, if a branch contained less than α nodes, it was discarded as a spur. The value of α was selected as 3, since all short branches were removed using this threshold value for our dataset. Figure 2.5d shows the sliced point cloud and Figure 2.5e shows its graphical representation after the skeleton graph pruning.

In the graph, every node has references to its neighbors, where the node is regarded as a parent if it has a smaller x value than its neighbor or otherwise, it is considered as a child. With this hierarchy, all the nodes can be categorized into three cases:

(1) Leaf tip (T): node with one neighbor.

(2) Internal node (I): node with two neighbors - usually has one child and one parent node.

(3) Junction (J): node with more than two neighbors - normally has only one parent and at least two child nodes.

It was observed that the junctions near the stem were always higher than the leaf collars because the slice near the leaf collar always contained both the stem and leaf base. To solve this problem, a skeleton refinement algorithm was implemented. Each branching node was split into two nodes by applying 2-means clustering to the point cluster associated with the original node. The splitting process continued until the distance between the two new nodes was less than a user-defined threshold. Based on trial and observation, we set the threshold to 0.02 m, which was approximately the stem diameter.

Step 3: Stem Identification

The stem in the merged point cloud could be approximately modeled as an elliptical cylinder (Chaivivatrakul et al., 2014; Frasson & Krajewski, 2010). However, if we directly searched for the points in the merged cloud that could be considered as a cylinder, the results always contained the part of rolling and straight young leaves (plant whorl). The whorl position was identified by searching for the junction node with largest x value in the skeleton (Figure 2.5e labels the node as *W*). To distinguish whorl from the shoot cylinder, the stem segmentation only considered those points lower than the plant whorl. After that, the 3D Hough transform (Dalitz et al., 2017) was applied to detect the central line of the cylinder model and the direction vector *v* of the line was calculated. The 3D Hough transform is an iterative voting process which selects the line with the most votes and removes the corresponding points after each iteration. Once the stem center line has been identified, the points corresponding to the shoot cylinder can be extracted by a thresholding-based strategy, i.e., the stem inliers are the points within a suitable radius from the central line (Figure 2.5e shows the stem center line in red and Figure 2.5f shows the stem inliers in orange).

Step 4: Leaf Segmentation

Leaf segmentation could be considered as a graph partitioning problem. We reconstructed each partial segment by iteratively traversing the graph from a leaf tip along a connected path of internal nodes until encountering a junction, the results of which can contain both leaves and tillers. Since sorghum leaves are very thin, the corresponding point cloud could be sparse if the camera captures the side view of a leaf, thus the filtering process described in 2.2.1 may cause disconnections of a single leaf. Before partial segmentation, a reconnection of the skeleton was employed to connect the broken leaves. The approach checked each one-degree node. If its distance to the nearest one-degree neighbor is less than the threshold (0.03m in our case, which is determined heuristically), then the two nodes were connected. Figure 2.5e shows the reconnected edge in green. Considering that leaves emerged from the shoot cylinder while tillers originated from the soil, we selected the partial segments connected with the stem as leaf candidates. Figure 2.5e shows that each individual leaf starts with a green node and ends with a red node. The stem and leaf segmentation is summarized in Algorithm 1.

Algorithm 1: Skeletonization algorithm for Stem and Leaf Segmentation

---

Input: C: The registered and filtered point cloud.

Output: Skeleton graph of the point cloud and determination of individual plant organs.

```
1: $P_{original}$ = segment(C);
    The region of interest, i.e., the sorghum plant, is extracted from the registered cloud by filtering out the points lower than the soil level.
2: $G_{initial}$ (V, E) = initialize - skeleton($P_{original}$);
    Compute the initial skeleton of the segment P and transform the 3D skeleton into a graph representation.
3: $G_{refined}$(V, E) = refine - skeleton ($G_{initial}$)
    Remove loops using Kruskal's algorithm and prune short branches using a threshold of $\alpha$ nodes.
4: {T, I, J} $\leftarrow$ V
    Categorize the set of nodes V into leaf tips (T), internal nodes (I) and junctions (J) based on the number of neighbors of each node.
5: W, $P_{part-plant}$ $\leftarrow$ detect - whorl ($G_{refined}$)
    Find the highest junction in J, label the junction node W as 'plant whorl', extract the points lower than the whorl as a cluster $P_1$, which will be used for stem segmentation.
6: $P_{stem}$ = stem - segmentation ($P_{part-plant}$)
    Detect the central line of the shoot cylinder in $P_{part-plant}$, identify the stem inliers ($P_{stem}$) from the data by extracting the points within a suitable radius from the central line.
7: $G_{reconnected}$ = skeleton - reconnection ($G_{refined}$)
```

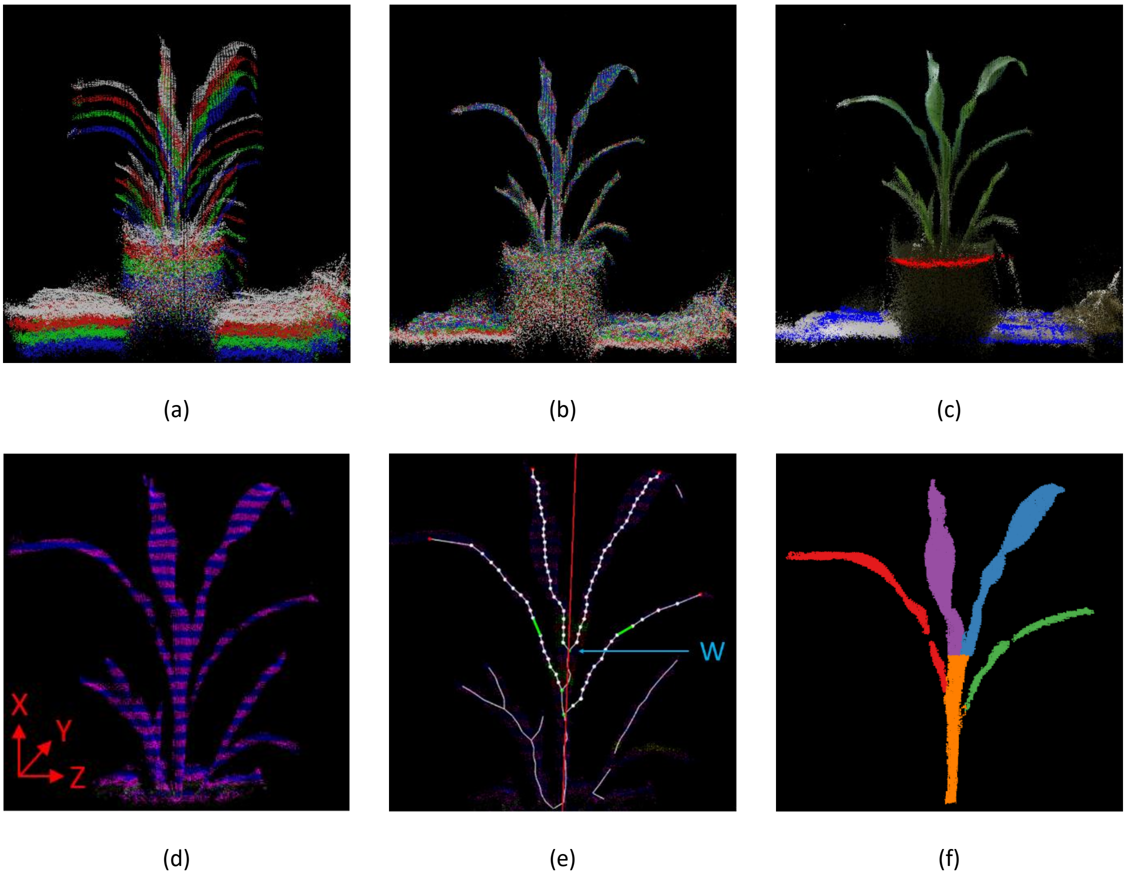
If the distance between a one-degree node and its nearest one-degree neighbor is $\leq \beta$, connect the two nodes.

8: $\{X_1, X_2, \ldots, X_n\}$ = partial - segment ($G_{reconnected}$)

Partition the skeleton graph into 1D partial segments such that each segment starts with T and ends with J. Each partial segment is defined as $X = \{N_1, N_2, \ldots, N_m\}$, where m is the number of nodes in the segment X.

9: $\{L_1, L_2, \ldots, L_n\} \in \{X_1, X_2, \ldots, X_n\}$

If one endpoint of a partial segment $X_i \in P_{stem}$, label the segment as a leaf (L_j).



(a) (b) (c)

(d) (e) (f)

Figure 2.5. Image processing pipeline: (a) Point clouds sampled from multiple heights. (b) Registered point cloud. (c) Ground and soil detection. Blue points are ground inliers and red points are soil inliers. (d) Sliced point cloud. (e) Hierarchical slice representation. The green point, red point, and white point represent junction, leaf tip, and internal node, respectively. The green lines are the reconnected skeleton of the broken leaves, and the red line denotes the stem central line. W stands for the positions of the junction node with the largest x value. (f) Stem and leaf segmentation. Different plant organs are visualized with different colors.

2.2.2.3. Morphological Traits Extraction

Five phenotypic parameters were extracted from the point clouds: plant height, stem diameter, stem volume, leaf angle, and leaf area. The multiple morphological traits obtained from the point clouds can be classified into two categories: (a) holistic traits and (b) organ-level traits. Holistic traits quantify the overall plant architecture, such as plant height and total leaf area. Organ-level traits analyze the individual organs of the plants, including stem diameter, leaf angle, and area of individual leaves. Figure 2.6 shows the definitions and principles of the measured phenotypes mentioned above.

Plant height: Plant height is defined as the distance from soil to the collar of the youngest fully expanded leaf, which is a useful and frequently measured trait in agronomic research (Neilson et al., 2015). In the graphical representation, the beginning of the stem was defined as the point where the stem emerges from the soil, and the end was defined as the second junction node from the top (where the corresponding leaf was typically fully expanded). The algorithm finds the second highest junction node (leaf collar) by checking the X coordinate of each node with those of two neighbors in the skeleton graph. Therefore, the plant height can be estimated as the distance from the node to the soil level along the stem direction using:

$$\mathrm{H} = \frac{\mathrm{X} - \mathrm{X}_{soil}}{cos\emptyset} \tag{2.4}$$

where $\emptyset$ is the angle between the stem center line and the Y-Z plane. X and X_{soil} denote the x-coordinate of the leaf collar and x-coordinate of the soil, respectively.

Stem diameter: Stem diameter was measured close to the soil to provide consistent and representative measurements of the entire stem. In the point cloud of the sorghum plant, the stem inliers immediately above the soil may not be complete due to occlusion by tillers, so eight slices of stem inliers whose distances to the soil ranged from 2cm to 10cm were used for stem diameter extraction (the thickness of each slice was 1 cm). The stem diameter was determined based on sorghum stem's roughly elliptical cross-section, hence the points of each slice were projected orthogonally onto the Y-Z plane (Figure 2.6a). Subsequently, ellipse fitting operation was applied to the projected points using an OpenCV library's build-in function (Bradski & Kaehler, 2008), which works by first finding the contour and then approximating an ellipse by minimizing the algebraic distance to its constraints (Fitzgibbon et al., 1999). The major axis of the fitted ellipse represents the stem width along the major axis of the elliptical cross-section (Figure 2.6b), therefore, the stem diameter can be calculated by averaging the major axes of all the ellipses of the stem slices.

Stem volume: Once the stem diameter has been extracted, the stem volume could be estimated based on the formula for the volume of an elliptical cylinder using:

$$\mathrm{V} = \frac{\pi \times D_a \times D_b \times \mathrm{H}}{4} \tag{2.5}$$

where H is the plant height, D_a and D_b denote the average major axis and minor axis of the fitted ellipses, respectively.

Leaf angle: Leaf angle is defined as the inclination between the leaf blade midrib and the stem. For leaf angle extraction, we used the leaf points within a radius of 0.06m (which was chosen empirically) from the central line of the stem cylinder model. To estimate the orientation of the leaf, the outer edge of the point cloud of the selected portion was first recognized (Figure 2.6c). Since the direction in which the leaves emerge was perpendicular to the sensor viewing direction in our experimental setup, the outer edge of the leaf seen by the depth sensor approximately represented the leaf blade midrib. Let $I(x, y, z)$ denote the edge points of the leaf. To recognize the orientation of the leaf, Principle Component Analysis (PCA) was performed on I. The PCA transformation computes a covariance matrix of I, which can be decomposed to a set of eigenvectors (u_1, u_2, u_3) and eigenvalues ($\lambda_1, \lambda_2, \lambda_3$). The eigenvector u_1 with the highest eigenvalue identifies a direction accounts for the greatest possible variance in the data set, hence providing an estimation of the overall orientation of the leaf. The Hough line fitting in stem segmentation process provided a direction vector (*v*) of the center line of the cylinder model of the stem. Then, the leaf angle was calculated as the angle between u_1 and v. An example of leaf angle measurement is shown in Figure 2.6, where white points are the edge points, and the yellow lines represent the direction of the major eigenvectors.

Leaf area and total area: To reconstruct the leaf surfaces, a greedy surface triangulation algorithm (Marton et al., 2009) was implemented to generate a triangle mesh based on a set of inliers of each individual leaf. The triangulation was performed based on projections of the local neighbors of a point along its normal. Figure 2.6d shows the reconstructed triangle mesh of a plant leaf. Leaf area was obtained by summing the triangle area of each leaf's surface mesh, and the total area included the surface area of all individual leaves and tillers.

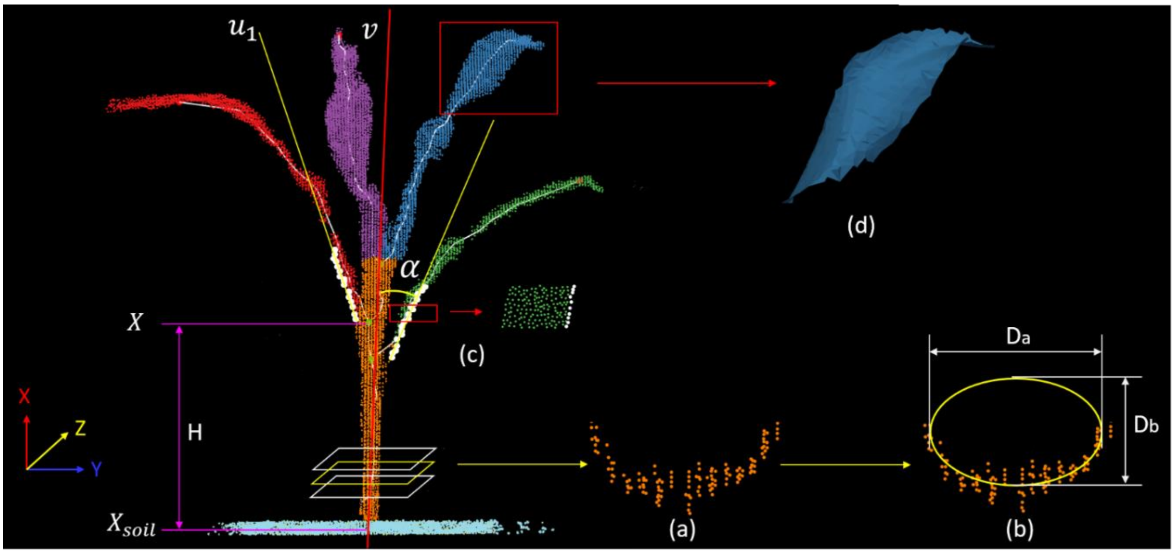


Figure 2.6. Phenotypic parameters: H - plant height; D_a - stem diameter (major axis of the fitted ellipse); D_b-minor axis of the fitted ellipse; α - leaf angle; (a) Orthogonal projection on Y-Z plane of the sliced stem point cloud; (b) Projected point cloud with ellipse fitting; (c) the outer edge of the leaf point cloud; (d) triangular mesh of leaf surfaces. In the figure, u1 and v are the direction vectors of the selected edge points and the stem cylinder, respectively. X and Xsoil denote the X coordinates of the leaf collar and the soil, respectively.

2.2.3. Performance Evaluation

Linear regression analyses were performed between the manual measurements and the system-derived measurements for all extracted parameters. The following statistics were calculated for performance evaluation: correlation coefficient (r) and root mean square error (RMSE, Equation 2.6).

$$\mathrm{RMSE} = \sqrt{mean(est - act)^2} \tag{2.6}$$

where *est* and *act* denote estimated values from the system and actual values from manual measurements, respectively.

The total area and stem volume obtained from the system were analyzed and compared with the wet and dry biomass using linear regression. The r value was calculated to evaluate the relationships between the actual parameters and those extracted from the system.

2.3. Results and Discussion

Figure 2.7 shows the 3D point cloud and the segmentation results of plant architectures for sorghum plants at different growth stages, where the individual leaf skeletons are visualized with random colors. The visualization of the 3D surface model appeared realistic, the result of the individual leaf segmentation matched the ground truth. It was noticed that the bottom leaf of the first plant in Figure 2.7c was missed, this was mainly because the too much overlap between the leaf and the tillers, as a result of which the skeletons of the leaf were removed as a part of the tillers. Overall, the results demonstrate that the segmentation algorithm was effective to extract each individual leaf from the 3D point cloud of the sorghum plants.

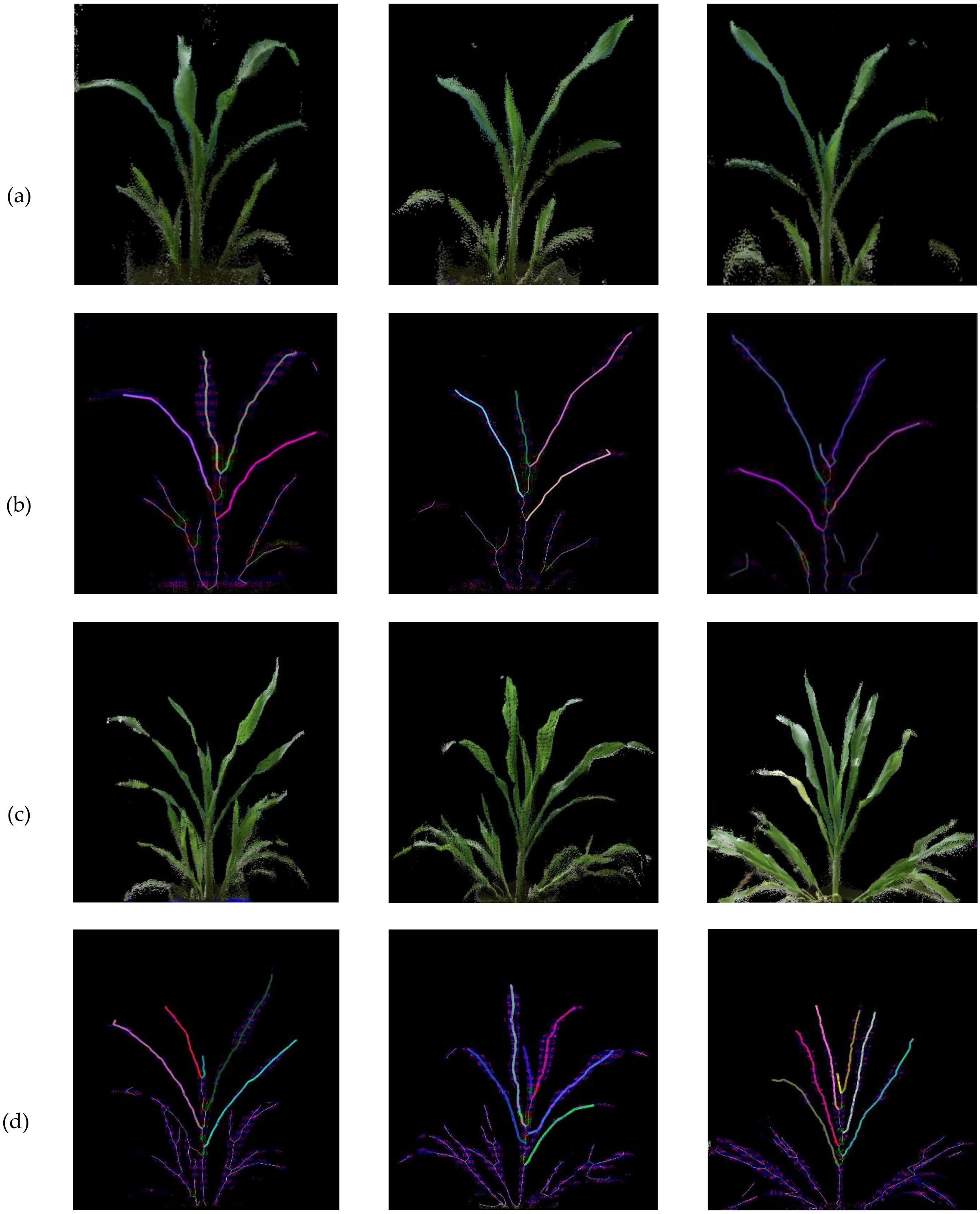


Figure 2.7. Point cloud and skeletons of sorghum plants at different growth stages. (a) point clouds of plants at W3. (c) point clouds of plants at W4 and W5. (b)(d) skeletons of plants with each leaf skeleton visualized with different colors.

2.3.1. Accuracy Assessment

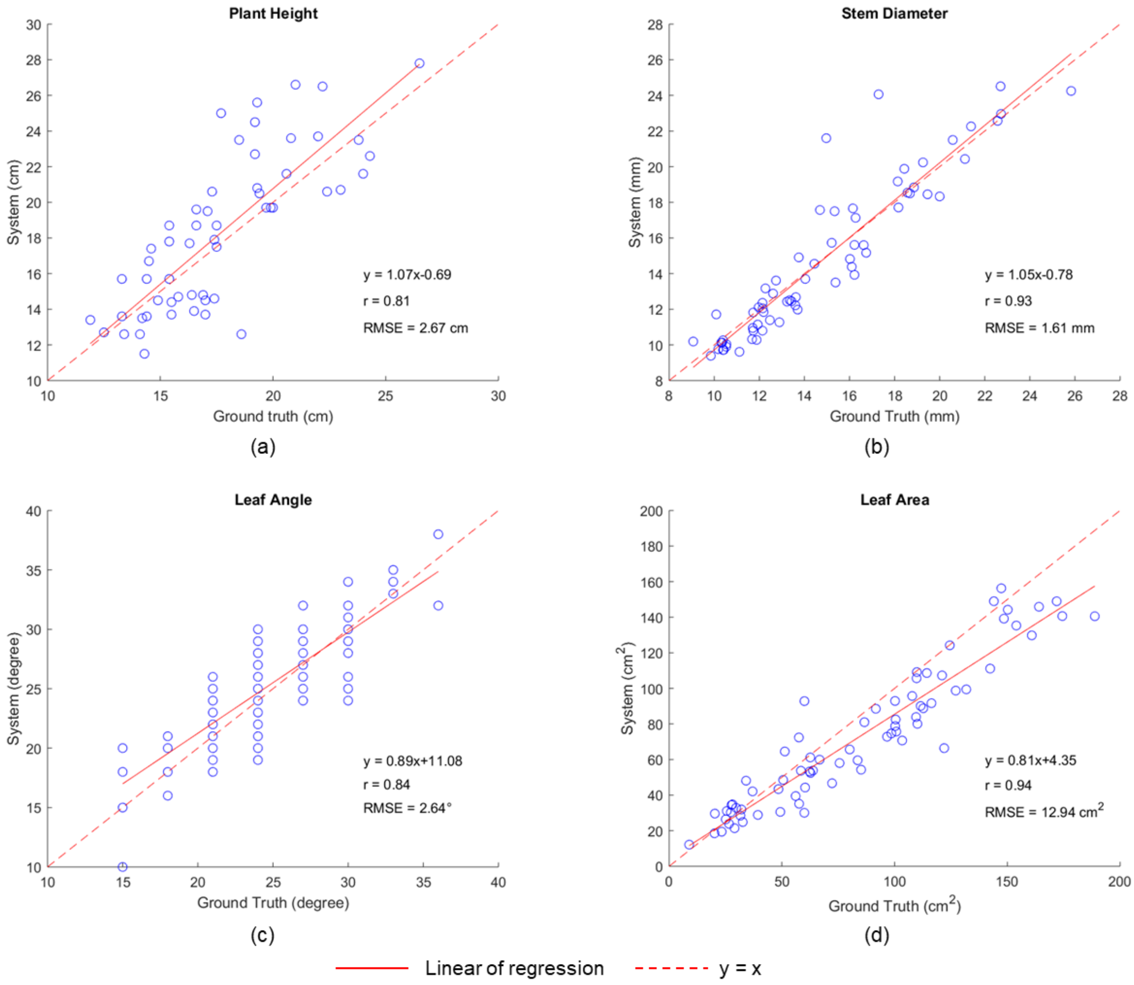


Figure 2.8. Linear regression results between the image-derived measurements and the manual measurements for plant height (a), stem diameter (b), leaf angle (c), and leaf area (d).

Comparisons of manual measurements and system-derived measurements of phenotypic parameters are shown in Figure 2.8. The results show that 3D point cloud analysis shows a promising potential for measuring the parameters inspected in this study, especially for stem diameter, leaf angle, and individual leaf area. The system-derived plant height is well correlated with the manual measurements (r = 0.81). A strong correlation (r = 0.93) was obtained for the stem diameter with an RMSE of 1.61 mm. The leaf angle yielded a coefficient of correlation of r = 0.84 with an RMSE of 2.64°. The image-derived leaf area was highly correlated with the manual measurement (r = 0.94), and the RMSE was 12.94 cm^2.

The plant height was manually measured using a tape from the soil to the collar of the last fully expanded leaf. However, the Kinect camera did not have the sufficient depth resolution to distinguish the leaf collars from other parts, thus making it challenging to detect collars in the point cloud. We defined the second highest junction as the leaf collar in our method, whereas for some sorghum plants, the youngest fully expanded leaves were located at the third highest junction, hence resulting in a relatively higher error for the plant height. A more precise way to find the collar of the last fully expanded leaf in the point cloud is needed to improve the accuracy of the plant height estimation. Furthermore, a more objective and clearer mathematical definition of the plant height needs to be investigated.

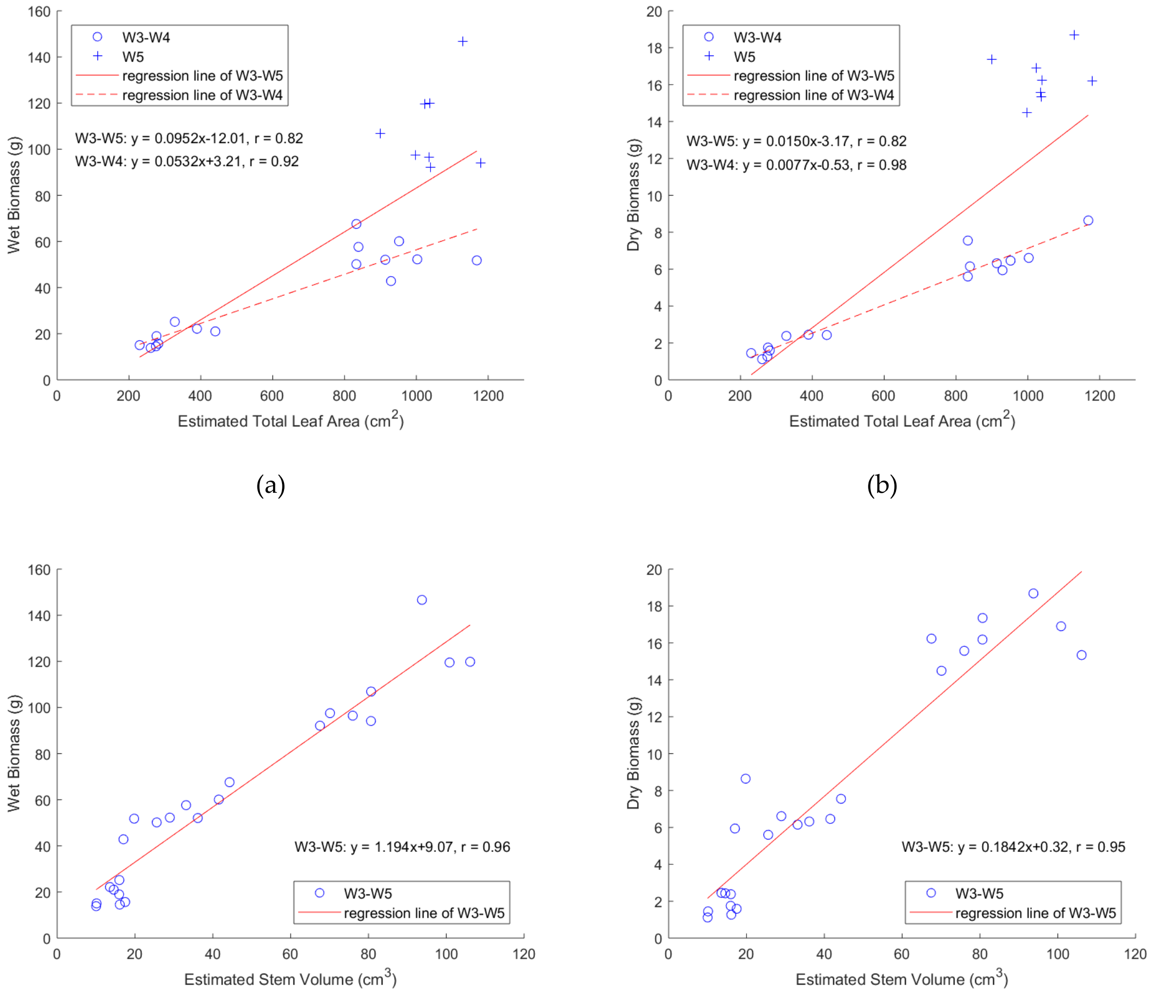
For stem diameter, the regression line of the data points matched well with the one-to-one line (Figure 2.8b), validating the point cloud processing method as an accurate technique for measuring sorghum stem diameter. The correlation coefficient (0.93) indicated a strong correlation between the estimated stem diameter of sorghum plants and the reference value. Given the average sorghum stem diameter was 14.78 mm, the RMSE (1.61 mm) presented an error rate of 10.9% of the average. It should be noticed that there were two outliers, which occurred because the tillers were located very close to the stem. Since the bottom slices were used for the stem cross-section calculations, the tiller proximity caused an overestimation of the stem diameter.

In spite of the narrow range of leaf angle observed in this experiment (15° ~ 36°), the RMSE was less than 3.00°, indicating that edge-based angle extraction via point cloud is well suited for determining the leaf angle. The random errors could be caused by two factors. First, the leaf angle extraction algorithm assumed that leaves were well aligned in the imaging plane, while in reality there existed some leaves distributed in other directions, resulting in that the outer edge points in the point cloud did not precisely denote the direction of the midrib. Second, some crooked and curled leaves did not have a smooth outer edge, which resulted in a large variance in edge point, leading to an inaccurate estimation of the leaf orientation with the use of the major eigenvector.

Overall, the leaf area derived from the system underestimated the actual leaf area, as most data points showed in Figure 2.8d were below the diagonal line. This is mainly caused by rolling and hidden parts of the upper leaves, since the Kinect sensor acquired point clouds only from the visible parts, resulting in a lack of information of leaves within the whorl. The leaf area was further underestimated during the process of triangulation, which tends to smooth the small variations in leaf surface and eliminate the local wrinkled structures particularly near the leaf tip and boundaries. This can be reduced by a more accurate triangulation methodology to reconstruct more details of leaf surfaces. Additionally, leaf occlusion and leaf rolling near the tips can also introduce errors. Nevertheless, the r value of 0.94 revealed that leaf area estimates were significantly correlated with the ground truth and the Kinect sensor can be an effective tool to measure leaf area.

2.3.2. Performance of Biomass Estimation

In addition to capturing morphological traits of plant architecture, the biomass prediction (fresh weight and dry weight) based on the extracted traits was investigated. Figure 2.9 showed the linear regression between the biomass and the system-derived traits. The total leaf area achieved similar results with wet and dry biomass from stage W3 to W5 (both r = 0.82). The stem volume had strong correlations with the wet biomass (r = 0.96) and dry biomass (r = 0.95). The values showed the promising potential of the Kinect v2 in biomass prediction.



(a) (b)

(c) (d)

Figure 2.9. Correlation of system-derived traits and biomass: (a) Estimated total leaf area and wet biomass. (b) Estimated total leaf area and dry biomass. (c) Estimated stem volume and wet biomass. (d) Estimated stem volume and dry biomass.

Though some previous studies reported the total leaf area had significant correlations with fresh weight and dry weight (Neilson et al., 2015; Yang et al., 2014; Zhang et al., 2017), the r values were not high when we using the data from stage W3 to W5. The variation in growth status among individual plants during W5 were larger than that during the other two growth periods (Figure 2.9a and Figure 2.9b). To address this problem, a linear regression was performed on the data in W3 and W4. When only the early stages were considered, the linear relationship between the total leaf area and the sorghum biomass showed noticeable improvement, with an r of 0.92 for the wet biomass and 0.98 for the dry biomass.

Compared with the total leaf area, the stem volume showed a significant improvement in correlation with the actual sorghum biomass, which added both stem height and stem size information to the model. Consistent with this, several studies have shown that biomass is related to stem height and diameter. Some researchers (George-Jaeggli et al., 2011) found that plant height can affect the biomass production of sorghum, and the reduced plant height was mainly due to a reduction in stem length to the collar of the flag leaf. Other studies have also shown that sorghum biomass has a positive relationship with stem diameter in both well-watered (Clough & Hunter, 2003) and drought-stressed conditions (Almodares et al., 2013).

For biomass (both fresh and dry weight) prediction, a noticeable improvement was achieved by adding stem-related features to the model. The stem volume generally achieved a stronger correlation with biomass than did the total leaf area. The total leaf area was more significant in predicting sorghum biomass at early stages (W3-W4) because of the more significant contribution of stem growth to biomass weight at later stages (W5). Therefore, it is possible to use Kinect v2 to estimate sorghum plant biomass, which can replace the intense labor needed to collect morphological traits related to biomass. Depth imaging is non-invasive to plants, and thus, it can be performed at different growth stages without damaging the plants.

2.4 Conclusions

We developed a low-cost machine vision system based on a commodity depth camera. The system was capable of vertically acquiring sequential side-view images of sorghum plants grown under controlled conditions at multiple developmental timepoints. A skeletonization algorithm was implemented to detect the individual leaves and distinguish the tillers by analyzing the 3D point cloud using a graph-based approach. The system was used to identify the sorghum plant architecture and characterize multiple important parameters. Automated measurements from the 3D surface model showed agreement with the corresponding manual measurements, validating the accuracy and utility of the system. In addition, both the total leaf area and stem volume showed potentials for wet and dry biomass prediction (r = 0.82 ~ 0.98). As such, depth imaging provides an efficient and economical solution for plant architecture characterization and phenotypic traits extraction, and hence can facilitate genomic studies and plant breeding programs. Future work will focus on improving the algorithms to address leaf occlusions better, and to adapt this approach to different growth environments and field scenarios.

Acknowledgements

This project was sponsored by the Plant Sciences Institute at Iowa State University. This material is based upon work supported by the National Science Foundation (Grant No. 1149603 to MGSF). MGSF was supported by the United States Department of Agriculture, National Institute of Food and Agriculture (Project #IOW04314). We thank Juan Panelo, Facundo Curin, and Ezequiel Delfino for their contributions to the imaged and manual phenotypic data collection.

Conflicts of Interest

The authors declare no conflict of interest.

References

Alenya, G., Dellen, B., & Torras, C. (2011). 3D modelling of leaves from color and ToF data for robotized plant measuring. *2011 IEEE International Conference on Robotics and Automation*, 3408–3414. https://doi.org/10.1109/ICRA.2011.5980092

Almodares, A. A., Hotjatabady, R. H., & Mirniam, E. (2013). Effects of drought stress on biomass and carbohydrate contents of two sweet sorghum cultivars. *Journal of Environmental Biology*, *34*(3), 585–589. http://www.ncbi.nlm.nih.gov/pubmed/24617146

An, N., Welch, S. M., Markelz, R. J. C., Baker, R. L., Palmer, C. M., Ta, J., Maloof, J. N., & Weinig, C. (2017). Quantifying time-series of leaf morphology using 2D and 3D photogrammetry methods for high-throughput plant phenotyping. *Computers and Electronics in Agriculture*, *135*, 222–232. https://doi.org/10.1016/j.compag.2017.02.001

Andújar, D., Dorado, J., Fernández-Quintanilla, C., Ribeiro, A., Andújar, D., Dorado, J., Fernández-Quintanilla, C., & Ribeiro, A. (2016). An Approach to the Use of Depth Cameras for Weed Volume Estimation. *Sensors*, *16*(7), 972. https://doi.org/10.3390/s16070972

Andújar, D., Ribeiro, A., Fernández-Quintanilla, C., & Dorado, J. (2016). Using depth cameras to extract structural parameters to assess the growth state and yield of cauliflower crops. *Computers and Electronics in Agriculture*, *122*, 67–73. https://doi.org/10.1016/J.COMPAG.2016.01.018

Apelt, F., Breuer, D., Nikoloski, Z., Stitt, M., & Kragler, F. (2015). *Phytotyping 4D : a light-field imaging system for non-invasive and accurate monitoring of spatio-temporal plant growth*. 693–706. https://doi.org/10.1111/tpj.12833

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

Bellasio, C., Olejníčková, J., Tesař, R., Šebela, D., Nedbal, L., Bellasio, C., Olejníčková, J., Tesař, R., Šebela, D., & Nedbal, L. (2012). Computer Reconstruction of Plant Growth and Chlorophyll Fluorescence Emission in Three Spatial Dimensions. *Sensors*, *12*(1), 1052–1071. https://doi.org/10.3390/s120101052

Borrell, A. K., Mullet, J. E., George-Jaeggli, B., van Oosterom, E. J., Hammer, G. L., Klein, P. E., & Jordan, D. R. (2014). Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. *Journal of Experimental Botany*, *65*(21), 6251–6263. https://doi.org/10.1093/jxb/eru232

Bradski, G. R., & Kaehler, A. (2008). *Learning OpenCV : computer vision with the OpenCV library* (1st ed.). O'Reilly.

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

Clough, A., & Hunter, M. N. (2003). Stem diameter: A rapid and accurate parameter for monitoring growth of sorghum. *Proceedings of the Australian Agronomy Conference, The Australian Society of Agronomy*. http://www.regional.org.au/au/asa/2003/p/4/clough.htm

Dalitz, C., Schramke, T., & Jeltsch, M. (2017). Iterative Hough Transform for Line Detection in 3D Point Clouds. *Image Processing On Line*, *7*, 184–196. https://doi.org/10.5201/ipol.2017.208

Dziubich, T., Szymański, J., Brzeski, A., Cychnerski, J., & Korłub, W. (2016). Depth Images Filtering In Distributed Streaming. *Polish Maritime Research*, *23*(2), 91–98. https://doi.org/10.1515/pomr-2016-0025

Eitel, J. U. H., Magney, T. S., Vierling, L. A., Brown, T. T., & Huggins, D. R. (2014). LiDAR based biomass and crop nitrogen estimates for rapid, non-destructive assessment of wheat nitrogen status. *Field Crops Research*, *159*, 21–32. https://doi.org/10.1016/J.FCR.2014.01.008

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395. https://doi.org/10.1145/358669.358692

Fitzgibbon, A., Pilu, M., & Fisher, R. B. (1999). Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(5), 476–480. https://doi.org/10.1109/34.765658

Frasson, R. P. de M., & Krajewski, W. F. (2010). Three-dimensional digital model of a maize plant. *Agricultural and Forest Meteorology*, *150*(3), 478–488. https://doi.org/10.1016/J.AGRFORMET.2010.01.003

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

Garrido, M., Paraforos, D., Reiser, D., Vázquez Arellano, M., Griepentrog, H., Valero, C., Garrido, M., Paraforos, D. S., Reiser, D., Vázquez Arellano, M., Griepentrog, H. W., & Valero, C. (2015). 3D Maize Plant Reconstruction Based on Georeferenced Overlapping LiDAR Point Clouds. *Remote Sensing*, *7*(12), 17077–17096. https://doi.org/10.3390/rs71215870

Gebbers, R., Ehlert, D., & Adamek, R. (2011). Rapid Mapping of the Leaf Area Index in Agricultural Crops. *Agronomy Journal*, *103*(5), 1532. https://doi.org/10.2134/agronj2011.0201

George-Jaeggli, B., Jordan, D. R., van Oosterom, E. J., & Hammer, G. L. (2011). Decrease in sorghum grain yield due to the dw3 dwarfing gene is caused by reduction in shoot biomass. *Field Crops Research*, *124*(2), 231–239. https://doi.org/10.1016/J.FCR.2011.07.005

Gibbs, J. A., Pound, M., French, A. P., Wells, D. M., Murchie, E., & Pridmore, T. (2018). Plant Phenotyping: An Active Vision Cell for Three-Dimensional Plant Shoot Reconstruction. *Plant Physiology*, *178*(2), 524–534. https://doi.org/10.1104/pp.18.00664

Holz, D., Ichim, A. E., Tombari, F., Rusu, R. B., & Behnke, S. (2015). Registration with the Point Cloud Library: A Modular Framework for Aligning in 3-D. *IEEE Robotics & Automation Magazine*, *22*(4), 110–124. https://doi.org/10.1109/MRA.2015.2432331

Hu, Y., Wang, L., Xiang, L., Wu, Q., Jiang, H., Hu, Y., Wang, L., Xiang, L., Wu, Q., & Jiang, H. (2018). Automatic Non-Destructive Growth Measurement of Leafy Vegetables Based on Kinect. *Sensors*, *18*(3), 806. https://doi.org/10.3390/s18030806

Jay, S., Rabatel, G., Hadoux, X., Moura, D., & Gorretta, N. (2015). In-field crop row phenotyping from 3D modeling performed using Structure from Motion. *Computers and Electronics in Agriculture*, *110*, 70–77. https://doi.org/10.1016/J.COMPAG.2014.09.021

Jiang, Y., Li, C., & Paterson, A. H. (2016). High throughput phenotyping of cotton plant height using depth images under field conditions. *Computers and Electronics in Agriculture*, *130*, 57–68. https://doi.org/10.1016/J.COMPAG.2016.09.017

Jiang, Y., Li, C., Paterson, A. H., Sun, S., Xu, R., & Robertson, J. (2018). Quantitative Analysis of Cotton Canopy Size in Field Conditions Using a Consumer-Grade RGB-D Camera. *Frontiers in Plant Science*, *8*, 2233. https://doi.org/10.3389/fpls.2017.02233

Kaminuma, E., Heida, N., Tsumoto, Y., Yamamoto, N., Goto, N., Okamoto, N., Konagaya, A., Matsui, M., & Toyoda, T. (2004). Automatic quantification of morphological traits via three-dimensional measurement of Arabidopsis. *Plant Journal*, *38*(2), 358–365. https://doi.org/10.1111/j.1365-313X.2004.02042.x

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, *7*(1), 48–48. https://doi.org/10.1090/S0002-9939-1956-0078686-7

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, L., Zhang, Q., Huang, D., 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

Lu, H., Tang, L., Whitham, S. A., & Mei, Y. (2017). A Robotic Platform for Corn Seedling Morphological Traits Characterization. *Sensors (Basel, Switzerland)*, *17*(9). https://doi.org/10.3390/s17092082

Marton, Z. C., Rusu, R. B., & Beetz, M. (2009). On fast surface reconstruction methods for large and noisy point clouds. *2009 IEEE International Conference on Robotics and Automation*, 3218–3223. https://doi.org/10.1109/ROBOT.2009.5152628

McCormick, R. F., Truong, S. K., & Mullet, J. E. (2016). 3D Sorghum Reconstructions from Depth Images Identify QTL Regulating Shoot Architecture. *Plant Physiology*, *172*(2), 823–834. https://doi.org/10.1104/pp.16.00948

Neilson, E. H., Edwards, A. M., Blomstedt, C. K., Berger, B., Møller, B. L., & Gleadow, R. M. (2015). Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time. *Journal of Experimental Botany*, *66*(7), 1817–1832. https://doi.org/10.1093/jxb/eru526

Nguyen, T., Slaughter, D., Max, N., Maloof, J., Sinha, N., Nguyen, T. T., Slaughter, D. C., Max, N., Maloof, J. N., & Sinha, N. (2015). Structured Light-Based 3D Reconstruction System for Plants. *Sensors*, *15*(8), 18587–18612. https://doi.org/10.3390/s150818587

Paulus, S., Behmann, J., Mahlein, A. K., Plümer, L., & Kuhlmann, H. (2014). Low-cost 3D systems: Suitable tools for plant phenotyping. *Sensors (Switzerland)*, *14*(2), 3001–3018. https://doi.org/10.3390/s140203001

Paulus, S., Dupuis, J., Mahlein, A.-K., & Kuhlmann, H. (2013). Surface feature based classification of plant organs from 3D laserscanned point clouds for plant phenotyping. *BMC Bioinformatics*, *14*(1), 238. https://doi.org/10.1186/1471-2105-14-238

Paulus, S., Dupuis, J., Riedel, S., Kuhlmann, H., Paulus, S., Dupuis, J., Riedel, S., & Kuhlmann, H. (2014). Automated Analysis of Barley Organs Using 3D Laser Scanning: An Approach for High Throughput Phenotyping. *Sensors*, *14*(7), 12670–12686. https://doi.org/10.3390/s140712670

Phillips, R. L. (2010). Mobilizing Science to Break Yield Barriers. *Crop Science*, *50*(Supplement_1), S-99. https://doi.org/10.2135/cropsci2009.09.0525

Pound, M. P., French, A. P., Murchie, E. H., & Pridmore, T. P. (2014). Automated recovery of three-dimensional models of plant shoots from multiple color images. *Plant Physiology*, *166*(4), 1688–1698. https://doi.org/10.1104/pp.114.248971

Rosell Polo, J. R., Sanz, R., Llorens, J., Arnó, J., Escolà, A., Ribes-Dasi, M., Masip, J., Camp, F., Gràcia, F., Solanelles, F., Pallejà, T., Val, L., Planas, S., Gil, E., & Palacín, J. (2009). A tractor-mounted scanning LIDAR for the non-destructive measurement of vegetative volume and surface area of tree-row plantations: A comparison with conventional destructive measurements. *Biosystems Engineering*, *102*(2), 128–134. https://doi.org/10.1016/J.BIOSYSTEMSENG.2008.10.009

Rovira-Más, F., Zhang, Q., & Reid, J. F. (2005). Creation of Three-dimensional Crop Maps based on Aerial Stereoimages. *Biosystems Engineering*, *90*(3), 251–259. https://doi.org/10.1016/J.BIOSYSTEMSENG.2004.11.013

Rusu, R. B., & Cousins, S. (2011). 3D is here: Point Cloud Library (PCL). *2011 IEEE International Conference on Robotics and Automation*, 1–4. https://doi.org/10.1109/ICRA.2011.5980567

Salas Fernandez, M. G., Bao, Y., Tang, L., & Schnable, P. S. (2017). A High-Throughput, Field-Based Phenotyping Technology for Tall Biomass Crops. *Plant Physiology*, *174*(4), 2008–2022. https://doi.org/10.1104/pp.17.00707

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

Sun, S., Li, C., Paterson, A., Sun, S., Li, C., & Paterson, A. H. (2017). In-Field High-Throughput Phenotyping of Cotton Plant Height Using LiDAR. *Remote Sensing*, *9*(4), 377. https://doi.org/10.3390/rs9040377

Tang, H., Brolly, M., Zhao, F., Strahler, A. H., Schaaf, C. L., Ganguly, S., Zhang, G., & Dubayah, R. (2014). Deriving and validating Leaf Area Index (LAI) at multiple spatial scales through lidar remote sensing: A case study in Sierra National Forest, CA. *Remote Sensing of Environment*, *143*, 131–141. https://doi.org/10.1016/J.RSE.2013.12.007

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 20260–20264. https://doi.org/10.1073/pnas.1116437108

White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., Feldmann, K. A., French, A. N., Heun, J. T., Hunsaker, D. J., Jenks, M. A., Kimball, B. A., Roth, R. L., Strand, R. J., Thorp, K. R., Wall, G. W., & Wang, G. (2012). Field-based phenomics for plant genetics research. *Field Crops Research*, *133*, 101–112. https://doi.org/10.1016/J.FCR.2012.04.003

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

Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., Fang, W., Feng, H., Xie, W., Lian, X., Wang, G., Luo, Q., Zhang, Q., Liu, Q., & Xiong, L. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nature Communications*, *5*(1), 5087. https://doi.org/10.1038/ncomms6087

Zhang, X., Huang, C., Wu, D., Qiao, F., Li, W., Duan, L., Wang, K., Xiao, Y., Chen, G., Liu, Q., Xiong, L., Yang, W., & Yan, J. (2017). High-Throughput Phenotyping and QTL Mapping Reveals the Genetic Architecture of Maize Plant Growth. *Plant Physiology*, *173*(3), 1554–1564. https://doi.org/10.1104/pp.16.01516

CHAPTER 3. MEASURING STEM DIAMETER OF SORGHUM PLANTS IN THE FIELD USING A HIGH-THROUGHPUT STEREO VISION SYSTEM

Modified from a manuscript published in *Transactions of the ASABE*

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

[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

Abstract

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 automate in the field due to the complexity of the imaging object and the environment. In recent years, stereo vision offers 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 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. 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 Region Convolutional Neural 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 root mean square error (RMSE) of 1.39 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.

3.1. 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 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 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’ (Mueller-Sim et al. 2017), ‘Vinobot’ (Shafiekhani et al. 2017), and some 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 like multispectral, 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., 2020), sorghum panicles (Malambo et al. 2019), and corn stems (Jin et al., 2019). However, such measurements can be complex and require long 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 (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 (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 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).

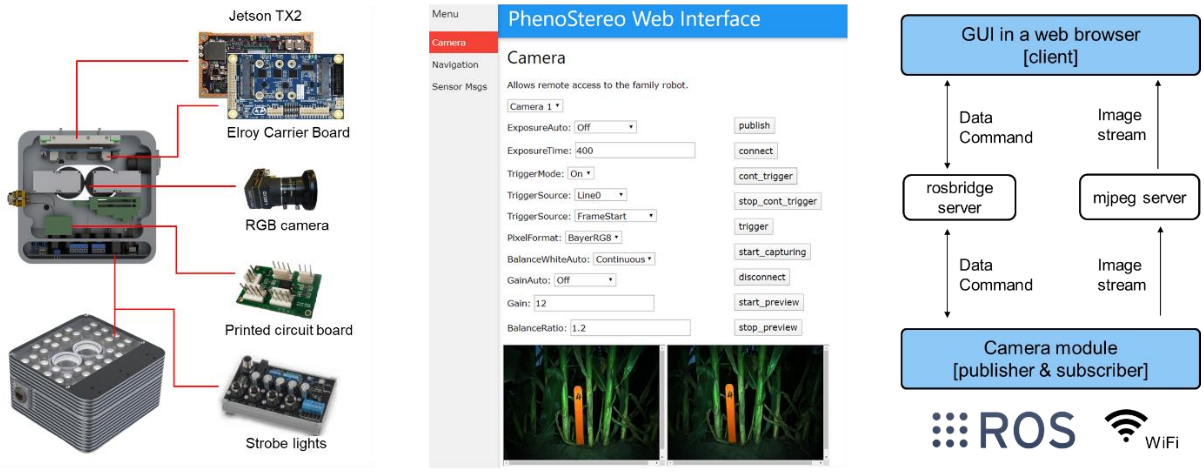
Sorghum is one of the most promising energy crops due to its tolerance to drought/heat stresses (Xin et al. 2008). Stem diameter is an important feature for yield estimation and stalk strength assessment of sorghum plants (Salas Fernandez et al., 2017). Typically, sorghum stem diameter is measured manually 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., 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 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. 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., 2020).

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) acquire high-quality images of sorghum plants in the field using our customized stereo module; 2) develop an automated image processing pipeline to detect the stems of sorghum plants and measure their stem diameter in reconstructed 3D models; 3) evaluate the performance of stem instance segmentation and stem diameter estimation.

3.2. Materials and Methods

3.2.1. 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. 3.1a) 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 and 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.



Allows remote access to the family robot.

Camera 1

ExposureAuto: Off

ExposureTime: 400

TriggerMode: On

TriggerSource: Line0

TriggerSource: FrameStart

PixelFormat: BayerRG8

BalanceWhiteAuto: Continuous

GainAuto: Off

Gain: 12

BalanceRatio: 1.2

(a) (b) (c)

Figure 3.1. The complete device. (a) Configuration of PhenoStereo imaging unit: consisting of stereo cameras and strobe lights for capturing high-resolution images. (b) The graphic user interface of PhenoStereo, which uses Robot Operating System (ROS) network for wireless control. (c) Overview of the software architecture.

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. 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 light illumination of 31000 lux at 0.5 m. A web-based graphic user interface (GUI) (fig. 3.1b) 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 software structure (fig. 3.1c) is composed two main modules: (i) the GUI that enables the control and the visualization, and (ii) the camera module that receive and execute 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 Robot Operating System (ROS) messages over 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 Rosbridge protocol for sending JSON-based messages and interacting with ROS, the other part is a collection of packages that includes a rosbridge server which provides a WebSocket connection for browsers and the camera (Coronado et al., 2020). To efficiently preview the images, the image stream was transferred in binary format using mjpeg sever. The control algorithm for image acquisition was written in C++, using ROS Kinetic running on Ubuntu 16.04.

3.2.2. Data Collection

A 4-wheel custom-built ground vehicle was used to deploy the PhenoStereo (fig. 3.2). The PhenoStereo was mounted on the vehicle at a height of 0.51 m (20 inches) above the ground, and aluminum frames were used to support the camera module. The ground vehicle was manually driven with a speed around 3 mph (4.8 km/h) through 30-inch (0.76 m) spaced crop rows and the images were captured at 10 FPS. A smartphone was used to run the web-based user interface to preview the images and adjust the camera parameters. The vehicle was running on a flat plot with a typical cultivated soil condition. 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 0.38 m (15 inches). Given the camera configuration and the field design, the system achieved a vertical field of view (FOV) of approximately 0.47 m and a horizontal FOV of 0.75 m.

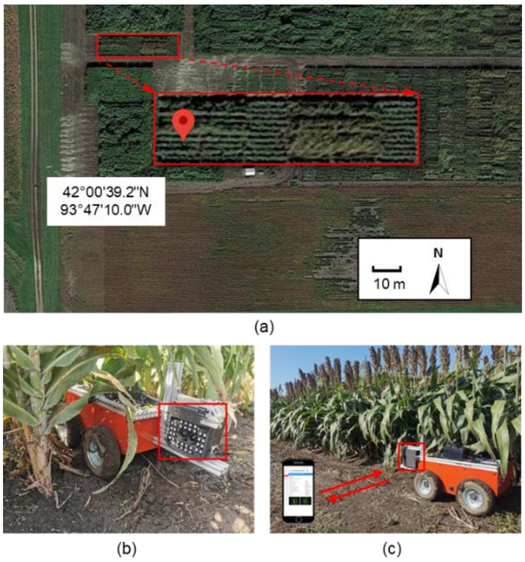


Figure 3.2. Data collection. (a) Location of the image collection. (b-c) Remote-controlled data collection with PhenoStereo under sunny (b) and overcast (c) weather conditions. A smartphone was used to interact with the PhenoStereo using wireless network connection.

Data were acquired at the Agricultural Engineering and Agronomy Research Farm of Iowa State University in Boone, Iowa (fig. 3.2a). 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 were collected under sunny (fig. 3.2b) and overcast (fig. 3.2c) weather conditions and at different times of day. 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. The measured plant was labeled with a tag to make it distinguishable during image processing.

3.2.3. 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. 3.3). 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.

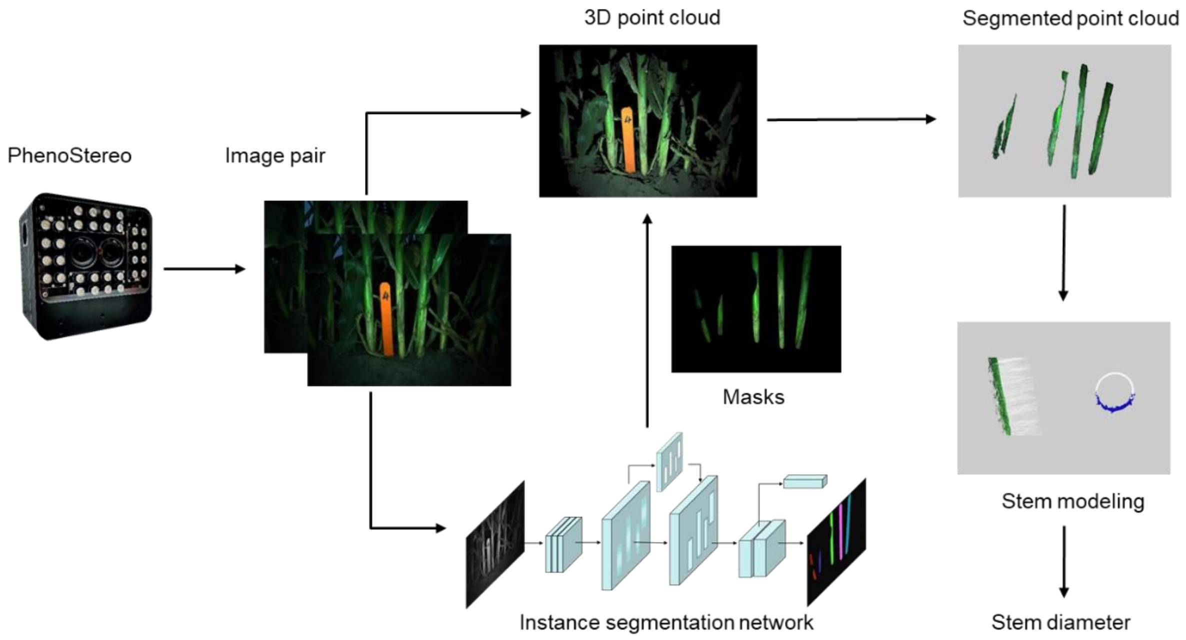


Figure 3.3. Illustration of the proposed image processing pipeline.

3.2.3.1 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 the triangulation principle and camera parameters.

The camera calibration was performed with a checkerboard pattern (Bradski & Kaehler, 2008) to compute the intrinsic and extrinsic parameters of stereo cameras. With the camera intrinsic parameters, the image pair was 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. 3.4a). The rectification process removes the lens distortion and enables the reduction of the search space of stereo correspondence from 2D to 1D.

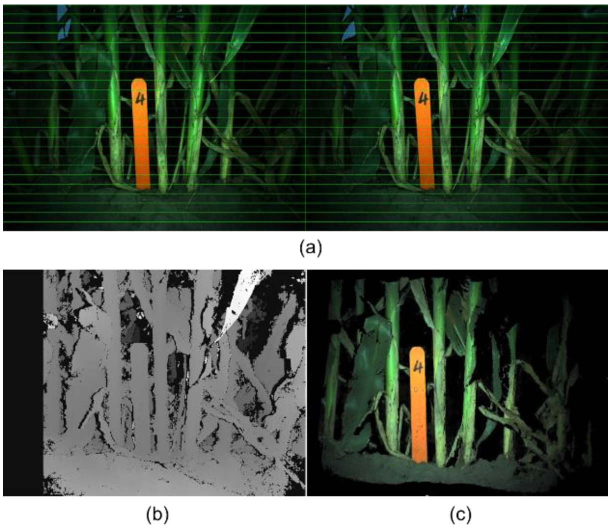


Figure 3.4. The procedure of stereo matching using Semi-Global Block Matching (SGBM). (a) A rectified and segmented image pair, where corresponding points are on the same horizontal scan lines. (b) The disparity output of the SGBM algorithm. (c) 3D point cloud reconstruction of the sorghum plant.

To extract the 3D coordinates of the object, the rectified image pair was used to generate a disparity map by stereo matching. 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. 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. Although this method is not as accurate as other advanced algorithms, it produced competitive results with better computational efficiency (Bao et al., 2019). In addition, SGBM showed a great potential in retaining edges and dealing with untextured areas (Hirschmüller, 2008), which is important for 3D reconstruction of in-field crop plants. 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 and 100 pixels, respectively. The result of stereo matching is the disparity map (fig. 3.4b), 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 3.1:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{x \times b}{d} \\ \frac{y \times b}{d} \\ \frac{f \times b}{d} \end{bmatrix} \tag{3.1}$$

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. 3.4c). 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.

3.2.3.2 Stem segmentation

For individual stalk size measurement, effective stem segmentation is the prerequisite for further image analysis. 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.

Mask R-CNN (He et al., 2020) was utilized to detect and segment sorghum stems from RGB images (fig. 3.5). The model first employs a feature pyramid network (FPN) (Lin et al., 2017) to construct feature maps of objects, 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 computational efficiency.

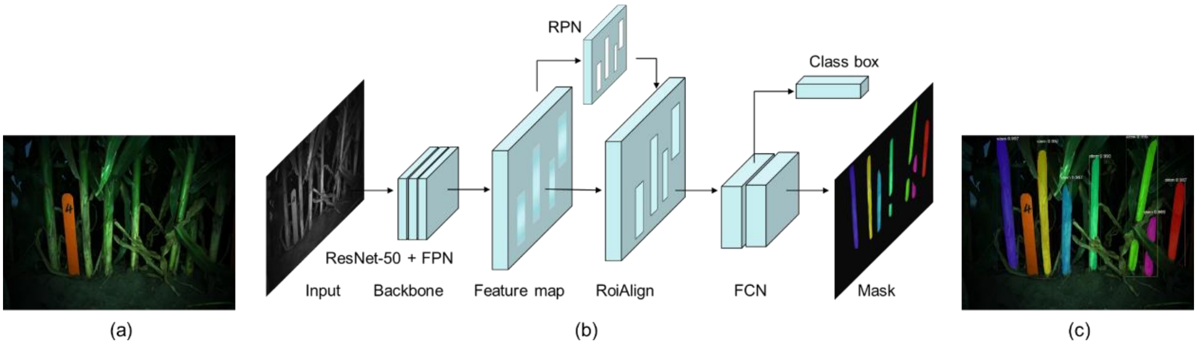


Figure 3.5. Summary of Mask R-CNN based stem segmentation. (a) An example of the input RGB image. (b) The diagram of Mask R-CNN architecture. (c) A masked image with detected stems.

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’. 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.

The Mask R-CNN algorithm was implemented on the TensorFlow framework. In this study, the ResNet network was used as the backbone for feature extractor due to its great potential in alleviating gradient disappearance and training degradation (Yu et al., 2019). Two ResNet backbone networks were experimented for the model, including ResNet-50 and ResNet-101. The model trained with ResNet-50 achieved better performance on the test dataset, therefore, ResNet-50 was used as the backbone network for stem segmentation. 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. 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. The 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 the segmentation of individual stems (fig. 3.5c).

3.2.3.3 Point cloud filtering

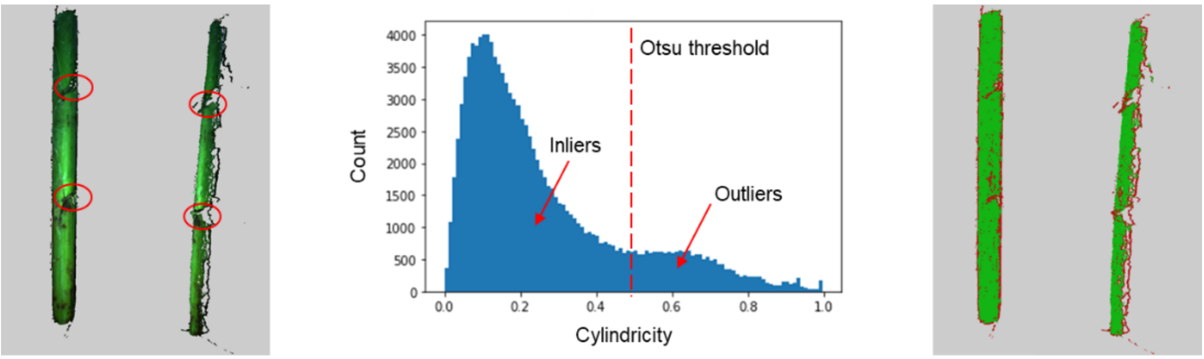
The filtering was one of the most important steps in the proposed algorithm since the accuracy of the cylinder fitting and stem diameter estimation depended on it. The process of filtering aimed to eliminate everything that does not belong to the cylindrical stem, including the random noise introduced 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. 3.6a), 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. 3.2), 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 3.3, where $\bar{p}$ (eq. 3.4) denotes the centroid of the region. We then performed PCA on C to obtain the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ $(\lambda_1 > \lambda_2 > \lambda_3)$.

$$\mathcal{N}_k = \{p_i : \|p - p_i\| < d\} \tag{3.2}$$

$$\mathrm{C} = \frac{1}{k}\sum_{i=1}^{k}(p_i - \bar{p}) \cdot (p_i - \bar{p})^T \tag{3.3}$$

$$\bar{p} = \frac{1}{k}\sum_{i=1}^{k} p_i \tag{3.4}$$



(a) (b) (c)

Figure 3.6. 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 3.5. 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. 3.6b), the points with cylindricity values greater than the Otsu threshold were removed as outliers (fig. 3.6c).

$$c = \frac{\lambda_1 - \lambda_2}{\lambda_1 - \lambda_3} \tag{3.5}$$

3.2.3.4 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. However, 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 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 perpendicular to the tangent plane at P (fig. 3.7a). 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 3.3. We calculated surface normal for each individual point in the point cloud (fig. 3.7b and c), and 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 3.6.

$$\mathcal{F}(a,b) = \sum_{i=1}^{m} (\vec{\omega} \cdot \vec{n_i})^2 = \sum_{i=1}^{m} (a \cdot x_i + b \cdot y_i + z_i)^2 \tag{3.6}$$

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.

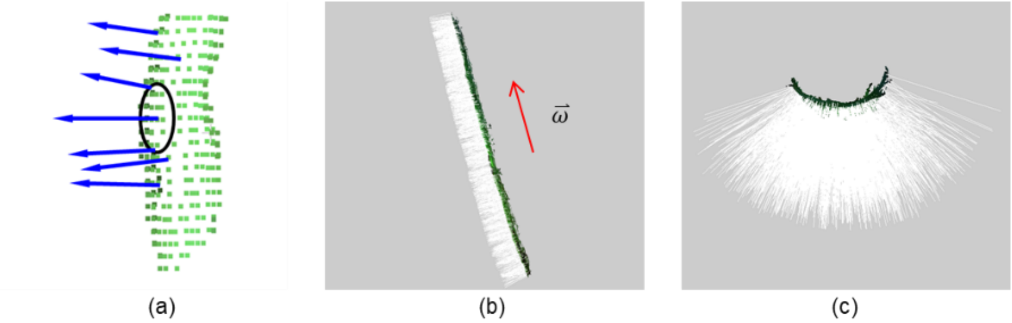


Figure 3.7. The estimation of the direction of the cylindrical axis. (a) Illustration of surface normal estimation. (b) The side view of the surface normals of a stem, the vector $\vec{\boldsymbol{\omega}}$ represents the direction of the cylindrical axis and the white epidermal hairs are the surface 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. 3.7) and the dot product (eq. 3.8) of the two vectors. According to Rodrigues' rotation formula, the rotation matrix (eq. 3.9) from vector $\overrightarrow{\omega_0}$ to vector $\overrightarrow{u_z}(0,0,1)$ can be obtained.

$$\vec{v} = \overrightarrow{\omega_0} \times \overrightarrow{u_z} \tag{3.7}$$

$$k = \overrightarrow{\omega_0} \cdot \overrightarrow{u_z} \tag{3.8}$$

$$R \; = I + [\sim\vec{v}] + [\sim\vec{v}]^2 \cdot \frac{1}{1+k} \tag{3.9}$$

where I is the identity matrix, and $[\sim\vec{v}]$ (eq. 3.10) is the skew-symmetric matrix of $\vec{v}$.

$$[\sim\vec{v}] = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix} \tag{3.10}$$

Finally, the stem point cloud was rotated by matrix R to ensure that the cylindrical axis was parallel to z-axis (fig. 3.8a). 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. 3.8b). 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. 3.11). The Levenberg-Marquardt algorithm (LMA) (Levenberg, 1944; Marquardt, 1963) was used for the optimization process.

$$\mathcal{F}(x_0, y_0, R) = \sum_{i=1}^{n} {d_i}^2 = \sum_{i=1}^{n} (\sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} - R)^2 \quad (3.11)$$

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.

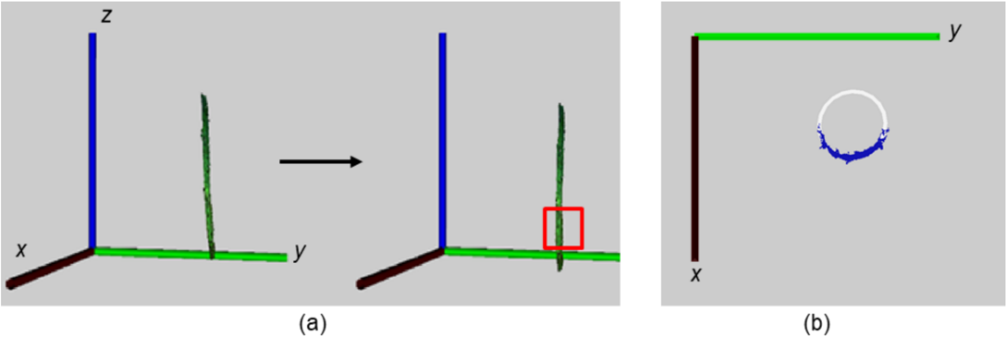


Figure 3.8. Circle detection using Levenberg-Marquardt algorithm (LMA). (a) 3D transformation of the stem point cloud, the point cloud was rotated to have its cylindrical axis to be parallel with z-axis. The stem section in the red square was selected for circle fitting. (b) Circle fitting in x-y plane, the blue points were the projected point cloud of the selected stem section, the white circle was the output of LMA.

3.2.4. Accuracy Assessment

To evaluate the performance of stem segmentation, three widely used evaluation metrics of precision (P), recall (R), and harmonic mean (F1) were chosen, which can be defined as:

$$P = \frac{TP}{TP + FP} \quad (3.13)$$

$$R = \frac{TP}{TP + FN} \quad (3.14)$$

$$F_1 = \frac{2 \cdot P \cdot R}{P + R} \tag{3.15}$$

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. In addition, the intersection over union (IoU) metric was used as an indicator for 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. 3.16), and mean absolute error (MAE, eq. 3.17).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N}(d_i - d_i^{gt})^2} \tag{3.16}$$

$$MAE = \frac{1}{N}\sum_{i}^{N}\left|d_i - d_i^{gt}\right| \tag{3.17}$$

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.

3.3. Results

3.3.1. Performance of Stem Segmentation

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 F1 over a variety of Intersection-over-Union (IoU) thresholds from 0.4 to 0.7 (Table 3.1). Considering IoU values equal to or lower than 0.5, all the precision, recall, and F1 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. Overall, the proposed Mask R-CNN segmentation approach produced satisfactory results for instance segmentation. Despite the relatively small size of training images, the model is adaptive to variable texture, stem size, occlusion, and illumination conditions.

Table 3.1. 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

Figure 3.9. 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.

3.3.2. Performance of Stem Diameter Estimation

The image-derived stem diameters were found to be highly correlated (r = 0.97) and accurate (RMSE = 1.39 mm) with respect to manually measured stem diameters (fig. 3.10). 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 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.

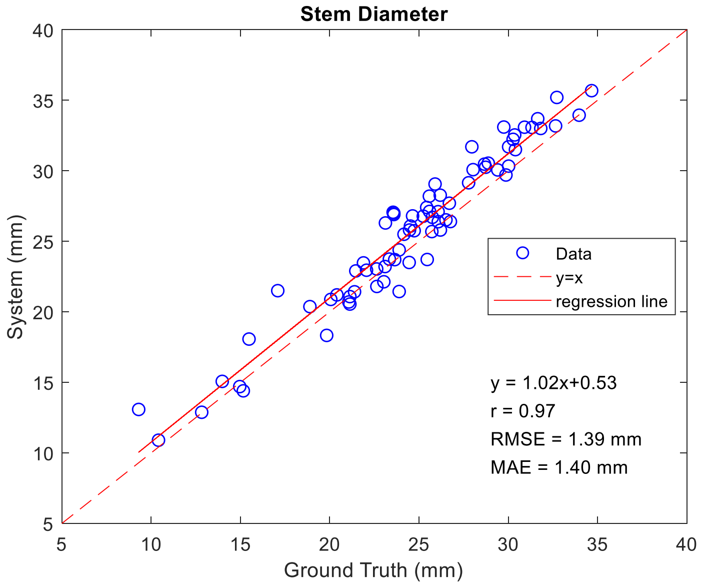


Figure 3.10. Correlation of system-derived stem diameter and ground truth

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.

3.3.3. 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 3.2. 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.

Table 3.2. 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

3.4. Discussion

3.4.1. Performance of PhenoStereo

In field conditions, the development of an imaging system that is robust to environmental conditions is challenging. The major challenges that come with field-based imaging include: 1) variable illumination, 2) wind, and 3) complex background. In this study, the active lighting-based camera system could potentially overcome the challenges of various lighting conditions. Meanwhile, the high-powered flash combined with faster shutter speed could reduce the adverse effects of motion due to wind. Furthermore, with dominating lighting on the viewed object, PhenoStereo is 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). With a customized strobe lighting system, PhenoStereo shows a great potential in overcoming these various conditions. 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.

3.4.2. 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. 3.11). 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 3.11. 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 & 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.

In this study, the disparity estimation was performed on both foreground and background, while 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 stage was the point cloud filtering mostly due to the expensive PCA analysis for each point. One possible solution to improve the efficiency could be implementing a down-sampling process to reduce the size of the point cloud before filtering.

3.5. 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 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 RMSE of 1.39 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.

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 Panelo 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. *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. *Field and Service Robotics. Springer Proceedings in Advanced Robotics,* 5, 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). 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 traitcharacterization 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

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), 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). Springer, Cham.

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

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, 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. *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

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, *The Plant Journal*, 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. 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

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

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 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

Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture*, *163*(April), 104846. https://doi.org/10.1016/j.compag.2019.06.001

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).

CHAPTER 4. FIELD-BASED ROBOTIC LEAF ANGLE DETECTION AND CHARACTERIZATION OF MAIZE PLANTS USING STEREO VISION AND DEEP CONVOLUTIONAL NEURAL NETWORKS

Modified from a manuscript submitted to *Journal of Field Robotics*

Lirong Xiang[a], Jingyao Gai[a,c], Yin Bao[a,d], Jianming Yu[b], Patrick S. Schnable[b], Lie Tang[a,*]

[a] Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011, USA

[b] Department of Agronomy, Iowa State University, Ames, IA 50011, USA

[c] College of Mechatronic Engineering, Guanxi University, Nanning 530007, China (current address)

[d] Department of Biosystems Engineering, Auburn University, Auburn, AL 36849, USA (current address)

Abstract

Leaf angle is an important architectural trait of crops due to its substantial role in light interception by the canopy and hence photosynthetic efficiency. Traditionally, leaf angle has been measured using a protractor, a process that is both slow and laborious. Efficiently measuring leaf angle under field conditions via imaging is challenging due to leaf density in the canopy and the resulting occlusions. However, advances in imaging technologies and machine learning have provided new tools for image acquisition and analysis that could be used to characterize leaf angle using three-dimensional (3D) models of field-grown plants. In this study, PhenoBot 3.0, a robotic vehicle designed to traverse between pairs of agronomically spaced rows of crops, was equipped with multiple tiers of PhenoStereo cameras to capture side-view images of maize plants in the field. PhenoStereo is a customized stereo camera module with integrated strobe lighting for high-speed stereoscopic image acquisition under variable outdoor lighting conditions. An automated image processing pipeline (AngleNet) was developed to measure leaf angles of nonoccluded leaves. This pipeline detects each leaf angle as a triplet of keypoints in two-dimensional images and extracts quantitative data from reconstructed 3D models. AngleNet-derived leaf angles and their associated internode heights were highly correlated with manually collected ground-truth measurements. Our study demonstrates the feasibility of using stereo vision to investigate the distribution of leaf angles in maize plant under field conditions. The proposed system is an efficient alternative to traditional leaf angle phenotyping that could accelerate breeding for improved plant architecture.

Keywords. Convolutional neural network, Field-based plant phenotyping, Leaf angle, Keypoint detection, Stereo vision

4.1. Introduction

Improving plant productivity and efficiency has become an important mission of plant breeding with the ever-increasing world population and climate change (Furbank and Tester, 2011). Understanding the adaptation of plants to various environments relies on dissecting the relationship between plant genotype (underlying genetic codes) and phenotype (e.g., plant architecture). Such knowledge can potentially aid in developing productive crop varieties and accelerate the plant breeding progress (Gibbs et al., 2018). Given the importance of maize as a cereal crop, plant breeders strive to generate high-yielding, stress-tolerant maize varieties (Che et al., 2020; Wang et al., 2015). Plant architecture plays an essential role in the interception of solar radiation (Duan et al., 2016; Truong et al., 2015). One component of plant architecture, leaf angle, has played an important part in crop improvement (Lewis et al., 2014; Tang et al., 2018). An optimal arrangement of leaves from the top to the bottom of the canopy can increase photosynthetic efficiency, thereby potentially increasing grain yield per unit area (Dzievit et al., 2019; Mantilla-Perez et al., 2020; Ort et al., 2015). Leaf angle is associated with increased productivity in maize (Hammer et al., 2009; Lewis et al., 2014; Mansfield & Mumm, 2014; Zhu et al., 2010). A canopy configuration with leaves at a more horizontal angle in the lower canopy and a more upright angle in the upper canopy is most desirable (Dzievit et al., 2019; Zhang et al., 2017).

Breeding for leaf angle requires measuring the leaf angles of large numbers of field-grown plants. The most common method is to select representative plants and manually measure the angles of typical leaves (Dzievit et al., 2019; Zhao et al., 2018). This slow process often fails to capture variation among leaf angles from the upper to lower portions of the crop canopy. Therefore, there is an urgent need to develop automated, high-throughput phenotyping methods for leaf angle (Mantilla-Perez et al., 2020; Zhang et al., 2017).

Advances in phenotyping platforms and sensing technologies provide an opportunity to improve the efficiency of leaf angle measurement (Duan et al., 2016; Mccormick et al., 2016). Several studies on methods for automated leaf angle measurements in maize plants using two-dimensional (2D) and three-dimensional (3D) images have recently been reported. For 2D imaging, RGB cameras are extensively used in indoor phenotyping systems to capture side-view images of maize plants (Cabrera-Bosquet et al., 2016; Kenchanmane Raju et al., 2020; Zhang et al., 2017). With controlled viewing angles, it is possible to derive accurate leaf angle measurements from 2D images (Kenchanmane Raju et al., 2020). Cabrera-Bosquet et al. (2016) used the PHENOARCH platform equipped with RGB cameras to capture 12 side-view images of each maize plant; the image with the most information was selected for leaf angle measurement. These 2D imaging-based methods have high throughput, but they require a specific viewing angle and are therefore not suitable for in-field leaf angle measurements. In contrast, 3D imaging that includes depth information provides the opportunity to overcome the limitations of 2D approaches (Apelt et al., 2015; Sun et al., 2020). Several technologies have been developed to reconstruct maize canopies and measure leaf angle in 3D space, such as 3D laser scanning (Wang et al., 2019), light detection and ranging (LiDAR) (Thapa et al., 2018), and time-of-flight (ToF) imaging (Chaivivatrakul et al., 2014). Wang et al. (2019) compared three representative 3D data acquisition approaches for maize phenotyping and obtained reliable measurements of leaf angle. However, all of these studies were performed on single potted maize plants under controlled imaging conditions.

With respect to field-based plant phenotyping, many platforms have been developed for field-based phenotyping in the last decade. The platforms can be categorized into aerial-based and ground-based. Aerial-based measuring systems have the advantages of high-throughput, portability, and maneuverability (Barbedo et al., 2019). However, the systems struggle to reveal phenotypic traits that are occluded from the top viewing angle. Ground-based phenotyping platforms, on the other hand, have the ability of traveling in the field, making it possible to capture side-view images and extract organ-level phenotypic traits such as leaf shape (Bao et al., 2019a), stalk size (Xiang et al., 2020), and fruit counts (Zabawa et al., 2020). Phenoliner (Kicherer et al., 2017; Zabawa et al., 2020) is a modified grapevine harvester with side-facing sensors for phenotyping of grape vines and berries. Such large high-clearance robot can support side-view proximal sensing but is not suitable for tall-growing plants such as corn and sorghum. In contrast, small robots that can traverse between crop rows are more flexible and portable. Robotanist (Mueller-Sim et al., 2017), a differential-steering robot developed for measuring bioenergy sorghum, is equipped with a vertical sensor rig and a robot manipulator for side-viewing phenotyping.

Developing field-based high-throughput phenotyping systems for agronomically-grown maize plants remains difficult because of the field and crop conditions including: (1) Narrow row spacing: agronomically spaced maize crop rows are typically 0.76 m (30 inch) apart in the US, and therefore, the robotic vehicle must have a narrow-body to traverse between crop rows and the imaging sensor must have a short working distance to acquire side-view images of maize plants; (2) Extreme plant height: Some maize plants can grow 10 ft tall and more, requiring multiple tiers of imaging sensors to cover the whole plant within the narrow agronomic row spacing; (3) Uneven ground surface: Running a narrow vehicle to image tall maize plants in a close vicinity on uneven field surfaces demands real-time balancing of the sensor mast; (4) Occlusion of plant canopies: There are serious occlusions of the plant organs (e.g., leaves, stalks) toward the imaging sensors due to either plant orientation or interferences from the neighboring plants, making 3D image analysis a necessity; (5) Environmental variations: variable lighting and wind conditions in the field can complicate the acquisition of high-quality images. Hence, the image sensor should have a high shutter speed and consistent lighting to overcome motion blur and variable lighting. Stereo vision with an active strobe lighting system provides a practical way for such purpose.

Various methods have been developed to detect and estimate leaf angle from images of maize plants under controlled environments. For example, Raju et al., (2020) used digital cameras to image greenhouse maize plants with controlled viewing angles. An image processing pipeline was developed to segment individual leaves by first binarizing the image using color information and then skeletonizing the plant using image thinning algorithms. Such binarization-skeletonization method is widely used in the leaf segmentation in 2D images (Das Choudhury et al., 2018; Kenchanmane Raju et al., 2020; Cabrera-Bosquet et al., 2016; Souza et al., 2021). However, the method requires controlled light conditions and viewing angles when taking images. The skeletonization approach has also been successfully used for interpreting the structure of 3D canopies, where a 3D skeleton was created by point cloud slicing along the growth direction (Bao et al., 2019a; Xiang et al., 2019; Zermas et al., 2020), or voxel thinning (Gaillard et al., 2020), or Laplacian contraction (Wu et al., 2019; Zhu et al., 2020). Skeleton play a crucial rule in analyzing the plant architecture and segmenting individual components, but the skeletonization process is highly dependent on spatial continuity and geometric features. Further, some methods need human interaction to achieve better leaf segmentation (Wu et al., 2019; Zhu et al., 2020). With comprehensive description provided by 3D models, clustering algorithms are often used in point cloud segmentation of maize plants (Atefi et al., 2019; Thapa et al., 2018). Thapa et al., (2018) implemented k-means clustering to separate individual leaves from point cloud after removing the stem and the background. However, the method is sensitive to noise and the user needs to specify the number of leaves in the algorithm. With the aim to avoid user interaction, some other clustering algorithms such as region growing, Euclidean clustering, and density-based spatial clustering of applications with noise (DBSCAN) were implemented for segmenting leaves in maize (Gélard et al., 2017) as well as vegetables (Hui et al., 2018; Xiao et al., 2020) and nursery crops (Li et al., 2019). Nevertheless, previous work has demonstrated that clustering methods are not feasible to segment close or overlapped leaves (Zhu et al., 2020). The existing methods for detecting and measuring leaf angle are more effective for single-potted maize plants with fully expanded and sparsely distributed leaves, but have limited ability to address the challenges in segmenting severely overlapped leaves in the field.

The leaf angle characterization of field-grown maize plants is regarded as one of the most challenging phenotyping tasks because of the substantial overlap and occlusion in maize plant canopies; and moreover, variable stem and leaf orientations relative to the camera can affect their visibility near a leaf collar (Bao et al., 2019b). To relieve these stresses, Wu et al. (2019) transported maize plants from the field to the greenhouse for 3D scanning, the resultant point cloud was first pre-processed manually and then skeletonized for leaf segmentation. As noted in the study, the 3D scanner is sensitive to wind, and even light wind caused by human moving would lead to overlaps and offsets in the point cloud. Additionally, the image acquisition process is fairly low-throughput and therefore not feasible for in-field applications. To measure maize plants in the field, Bao et al. (2019b) developed an automated system to characterize plant architectural traits based on ToF 3D imaging. Plants were detected as 3D Hough lines and a skeletonization algorithm was developed to separate stems and leaves. However, the estimation of leaf angle suffered from occlusion and overlapped canopies especially for the plants at late growth stages. Additionally, this method requires increased inter-row and intra-row spacing to reduce occlusions and to achieve a sufficient field of view for the depth camera, and the images must be acquired near or after sunset due to the ToF camera's susceptibility to sunlight. Due to the dense canopy and severe occlusion, especially during later periods of plant growth, the segmentation of individual plants and leaves can be difficult when plants are grown at agronomic field densities. Therefore, several plot-level architectural traits such as plot-based plant width (PPW) (Mantilla-Perez et al., 2020) have been developed to estimate leaf angle indirectly. However, such plot-level descriptors can be influenced by other architectural traits: for example, longer leaves with the same leaf angles may result in a larger PPW value. This method of leaf angle measurement was seriously compromised because the severe occlusions generated by the dense plant population and dense canopy of sorghum plants have made automated leaf angle detection largely impossible. Following recent breakthroughs, deep convolutional neural networks (CNNs) have strong performance on image processing tasks in field-based plant phenotyping studies (Pérez-Borrero et al., 2020; Santos et al., 2020; Zou et al., 2020). Compared with traditional computer vision methods, CNN-based approaches can better cope with the image-by-image variations caused by the differences in occlusion, illumination, and viewpoints, hence providing new opportunities for automation (Koirala et al., 2019; Vit et al., 2020). The morphological trait characterization step in plant phenotyping can be treated as an object detection problem, in which CNN models can be trained to find the regions or points of interest (Jiang & Li, 2020). For example, in a study measuring stem diameter in sorghum plants in the field, a region-based CNN model was used to detect sorghum stems in a given image (Xiang et al., 2020). In another study, this model was used to identify points for length phenotyping (Vit et al., 2020). Since leaf angle is defined as the angle made by the midrib of a leaf blade and a stem, the regions near leaf collars are the most important (Bao et al., 2019b; Kenchanmane Raju et al., 2020). Furthermore, in plants with known architectures, the topology of a leaf angle can be defined by a triplet of keypoints, including a point on the midrib (M), a point on the stem (S), and a point near the leaf collar (C). Keypoint detection is widely used for human pose estimation and facial recognition, but its use in plant phenotyping is less common. Based on the success of object and keypoint detection, it is worth exploring how CNN-based approaches can be adapted to detect the distinct regions of leaf collars and to identify the points of interest used to quantify leaf angles.

In this study, we focused on the problems associated with detecting and measuring leaf angles along the entire height of maize plants based on 3D models reconstructed from stereoscopic images. The specific objectives were to (1) develop a customized stereo-vision-based plant phenotyping platform to image maize plant canopy at different heights in the field; (2) develop an automated image processing pipeline to detect the regions and keypoints of interest and characterize plant architecture via a 3D modeling process and derive two important traits: leaf angle and its associated node height; (3) evaluate the performance of the proposed approach by comparing system-derived measurements with ground truth; and (4) explore the effectiveness of the newly developed system for characterizing leaf angle variations in different maize inbred lines in the field.

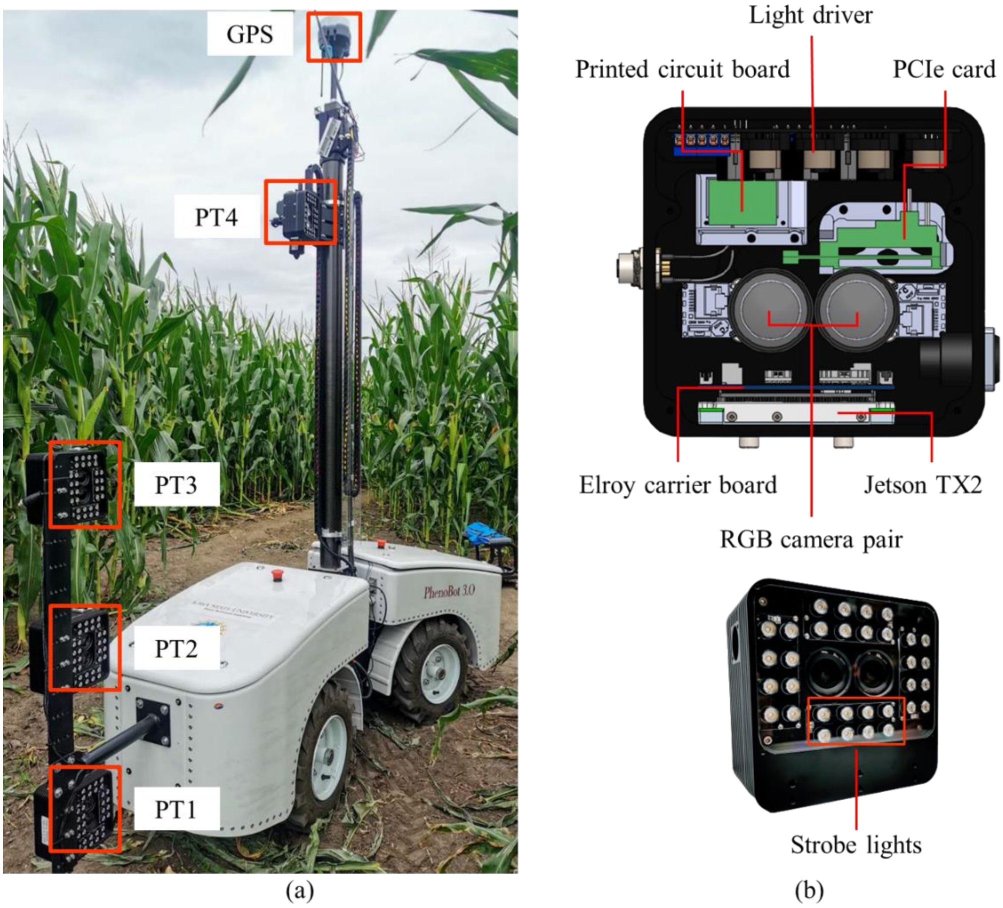
4.2. Methods

4.2.1. PhenoBot 3.0 and PhenoStereo

PhenoBot 3.0 (Tuel, 2019) (Figure 4.1a) is a robotic ground vehicle designed for field-based plant phenotyping, especially for tall-growing plants such as maize and sorghum. With a unique narrow body design (20-inch or 0.508-m width), this robot is able to navigate between crop rows with 76-cm spacing. Additionally, the robot featured a centrally articulated steering mechanism, which is more efficient and space saving to operate in narrow spaces. A sensor mast with adjustable height (between 2.1 m and 3.7 m) is mounted on the vehicle to support an RTK-GPS module and images sensors. To avoid collision with the crop rows and keep a workable camera-to-object distance, the roll angle of the sensor mast was actively controlled to maintain the mast gravitationally vertical in the presence of uneven ground surfaces. The self-balanced sensor mast enables the phenotyping platform to image plant sections at different heights to cover the whole canopy. The whole system is designed to be self-navigated with the integration of various sensors including an RTK-GPS module, a ToF camera, and two Inertia measurement units (Gai, 2020). In the case of GPS signals and radio communication signals are hindered by dense canopy, a ToF camera at the front of the vehicle is used for local navigation by detecting parallel rows in depth images. With the self-balanced sensor mast and the image sensor configuration, PhenoBot 3.0 is capable of capturing various organ-level phenotypic traits located at different heights of maize plant canopy, such as leaf angle, stem diameter, ear height, tassel size, etc.

PhenoStereo (Xiang et al., 2020) is a custom-built stereo camera module for field-based phenotypic data acquisition (Figure 4.1b). The module houses two industrial color cameras (Phoenix 3.2 MP, Lucid Vision Labs, Richmond, B.C., Canada) with spatial resolution of 1536×2048 pixels. Each color camera is equipped with a lens of 4.0 mm focal length. To overcome the variable outdoor lighting conditions, the module integrates an embedded computer and a strobe lighting system for high-speed stereoscopic image capturing at a rate up to 14 frames/s (FPS). The active lighting system uses high-intensity LEDs (LM75, Smart Vision Lights, Norton Shores, Mich.) equipped with a 576 W (with 10% duty cycle) driver (CTL-IO-4, Smart Vision Lights). In all experiments, the aperture of the lens and the strobe time were kept fixed at f/11 and 1.5 ms, respectively. With a short baseline

(0.038 m) and wide angles of view (85.8°V × 63.6°H), PhenoStereo has a short working distance and is suitable for collecting image data in between agronomically spaced crop rows. PhenoBot 3.0 is equipped with four tiers of PhenoStereo cameras to capture side-view and close-range images of maize plants in the field (Figure 4.1a). The four cameras used in this study (from bottom to top) were referred to as PT1 to PT4. One camera (PT4) was mounted on the sensor mast to photograph the canopy near the maize tassel, and the other three were mounted on a customized frame in the back of the vehicle. Given the 30” row spacing and the camera specifications, the system configuration resulted in a coverage of 2.22 m of the canopy in the vertical direction.



(a)

(b)

Figure 4.1. PhenoBot 3.0 and PhenoStereo. (a) PhenoBot 3.0 with a self-balanced sensor mast. PhenoStereo cameras were mounted at four different heights to capture side-view images of maize plants. (b) PhenoStereo with a self-contained embedded design and integrated strobe lights.

The software control system (Figure 4.2) was implemented in Robot Operation System (ROS) (Stanford Artificial Intelligence Laboratory et al., 2018). Each PhenoStereo camera was a ROS node, and a *roscore* node was run on the computer enclosed in PT1. A graphic user interface (GUI) based on *Robot Web Tools* (Toris et al., 2015) was developed to serve inputs and outputs. During data collection, the camera parameters and control commands were sent through the GUI and interacted with the ROS nodes wirelessly through the *rosbridge* protocol (Crick et al., 2017). In addition, the image stream was published by the ROS nodes and displayed on the GUI through the *mjpeg server* to adjust camera parameters and preview images.

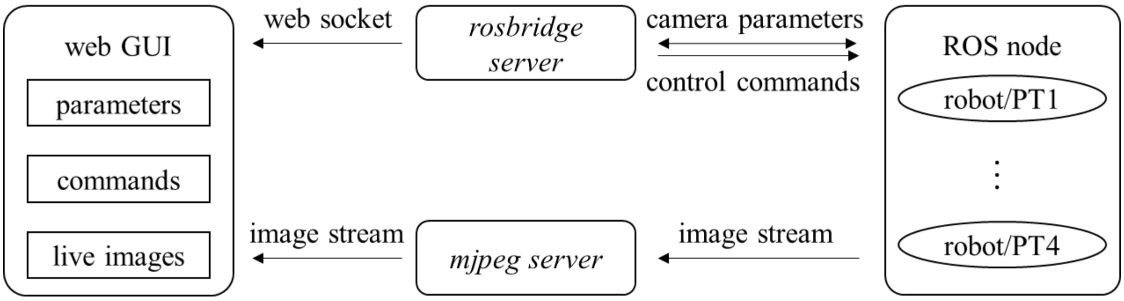


Figure 4.2. Overview of the software architecture.

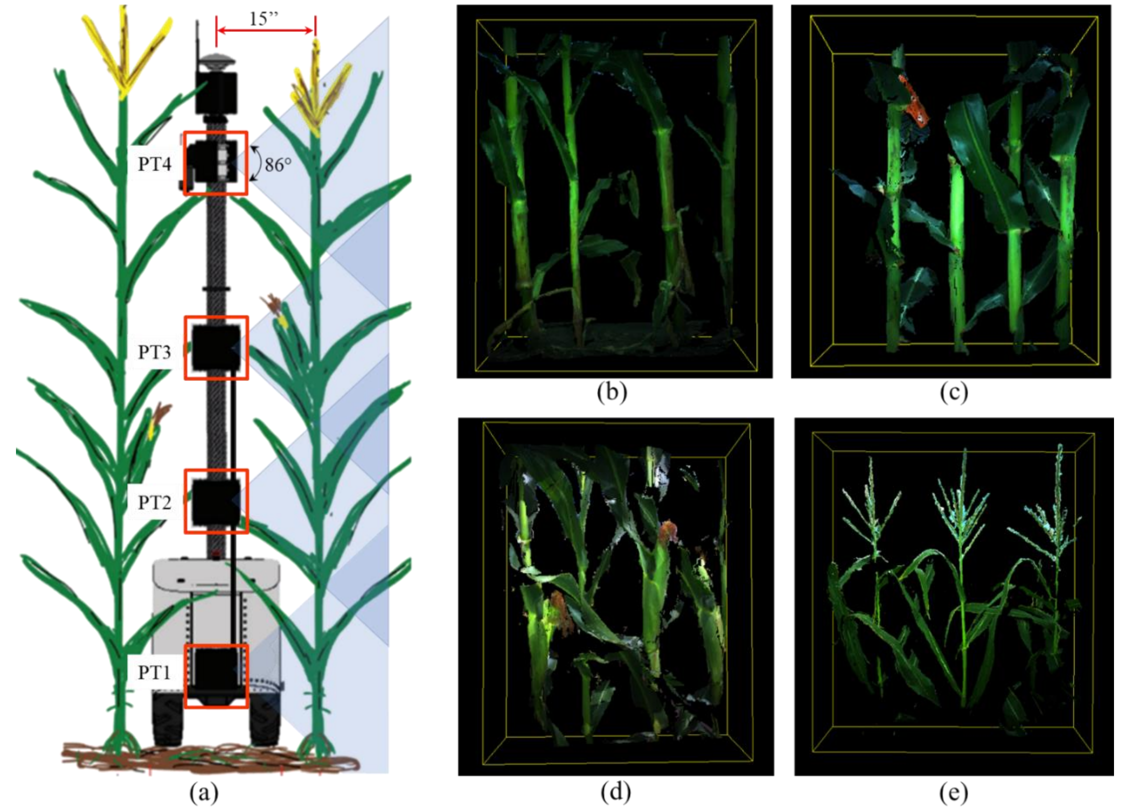
4.2.2. Data Acquisition

The field image acquisition was carried out at two test sites with different environmental conditions and multiple maize inbred lines. The data sets for system validation including (1) PS-Boone: PhenoBot acquired with four sets of PhenoStereo cameras at the Agricultural Engineering and Agronomy Research Farm of Iowa State University; and (2) PS-Ames: PhenoBot acquired with three sets of PhenoStereo cameras at the Curtiss Farm of Iowa State University. A summary of the data acquisition details is presented in Table 4.1. Two inbred lines, B73 (Russell, 1972) and Mo17 (Zuber, 1973), were used to investigate leaf angle architecture, and the remaining lines were randomly selected and used for system validation. Image data were collected on sunny days without the use of a shading structure. The mounting positions of PhenoStereo cameras were adjusted to cover the entire canopy.

During data collection, Phenobot 3.0 navigated between the crop rows with the PhenoStereo cameras capturing side-view images of plants (Figure 4.3a). The shutter speed was set to 0.3 ms to avoid motion blur induced by robot motion, wind, and robot vibration. Images at different heights were collected that in aggregate covered most or all of the maize canopy (Figures 4.3b–e).

Table 4.1. The characteristics of the two datasets used for system validation

Dataset	PS4-Boone	PS3-Ames
Location	Agronomy Farm, Boone, IA	Curtiss Farm, Ames, IA
Operation time	1-3 p.m., August 5, 2020	10-11 a.m., August 6, 2021
Crop row spacing	0.76 m	3.35 m
Number of plants measured	10	6
Leaf angle range	15-50º	5-60º
Node height range	0.3-2.0 m	0.2-1.6 m
Camera-to-object distance	0.38-0.51 m	0.51-0.64 m
Frame rate	5 FPS	10 FPS
Driving speed	~1 m/s	~2 m/s
Weather conditions	bright sunlight, light wind	light sunlight, moderate wind

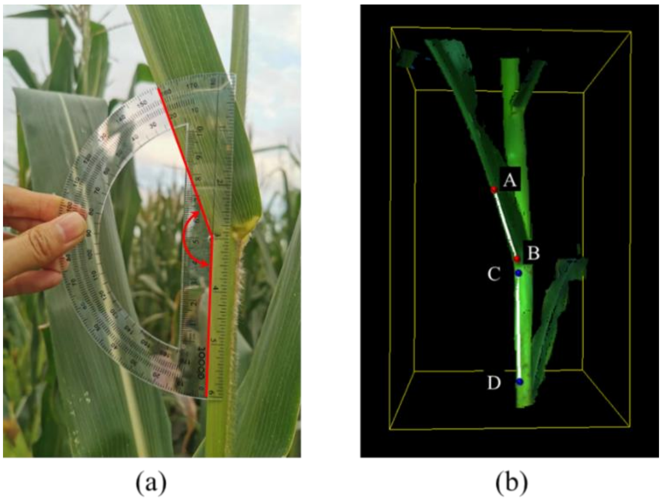


(a) (b) (c) (d) (e)

Figure 4.3. Data collection and visualization. (a) PhenoBot 3.0 with the four-level stereo imaging configuration traversing between crop rows with 0.76-m (30-inch) spacing to capture

side-view images. (b–e) The plant canopy imaged at different heights. Point clouds (b), (c), (d), and (e) are reconstructed 3D models from PT1, PT2, PT3, and PT4, respectively.

We measured the angle between the midrib of a leaf and the stem segment below the leaf collar (Figure 4.4a); the supplementary angle of the measured angle served as the leaf angle. The associated node height of a leaf angle was defined as the vertical distance from the ground to the leaf collar. The ground-truth data for leaf angle were collected both from manual measurements in the field and independently from the reconstructed 3D models. For field measurements, leaf angles were measured using a protractor with its origin placed on the leaf collar (Figure 4.4a). To measure leaf angles from 3D models, a program was developed in which a user clicks four points (A, B, C, and D) on the point cloud to identify the direction of the stem and the leaf (Figure 4.4b). Node height ground-truth data were obtained using a tape measure in the field.



(a) (b)

Figure 4.4. Two types of ground-truth measurements for leaf angle. (a) Manually measuring leaf angle in the field using a protractor. (b) Manually measuring leaf angle in the 3D point cloud.

4.2.3. Image Processing

In this study, we developed a unique image processing pipeline named AngleNet for leaf angle detection and characterization (Figure 4.5). This pipeline consists of three major steps. First, a CNN-based model is trained to detect regions around leaf collars (Figure 4.5a) and three keypoints in each region: C, M and S (Figure 4.5b). The two line segments (CS and

CM) formed by the three keypoints play an important role in the leaf angle extraction step. The detection is performed on the right image. To improve the computational efficiency, a cropped image pair derived from left and right images based on the detected region of interest (RoI) is then used for stereo matching, and a disparity map is generated for 3D reconstruction (Figure 4.5c). The coordinate system is defined as follows: the *x* axis is parallel to the robotic vehicle's heading direction, the *y* axis is the vertical direction of plant growth, and the *z* axis is perpendicular to the *x*–*y* plane pointing towards the plants. In this study, (*x*, *y*) represents the 2D image point coordinates in pixels, and (*x*, *y*, *z*) represents the 3D coordinates of a 2D point. To include the detected bounding boxes and ensure overlap between the left and right images, the cropped region for the image pair is centered at (*x*, *y*) with 2*W width and 2*H height, where (*x*, *y*) is the center of the bounding box detected in the right image, W and H are the width and height of the bounding box, respectively. Finally, the detection results in the 2D images are reprojected onto their 3D coordinates based on stereo camera calibration, and the random sample consensus (RANSAC) algorithm is used to fit two 3D lines to quantify the leaf angle (Figure 4.5d). At the same time, the node height is calculated based on the *y* coordinate of the leaf collar.

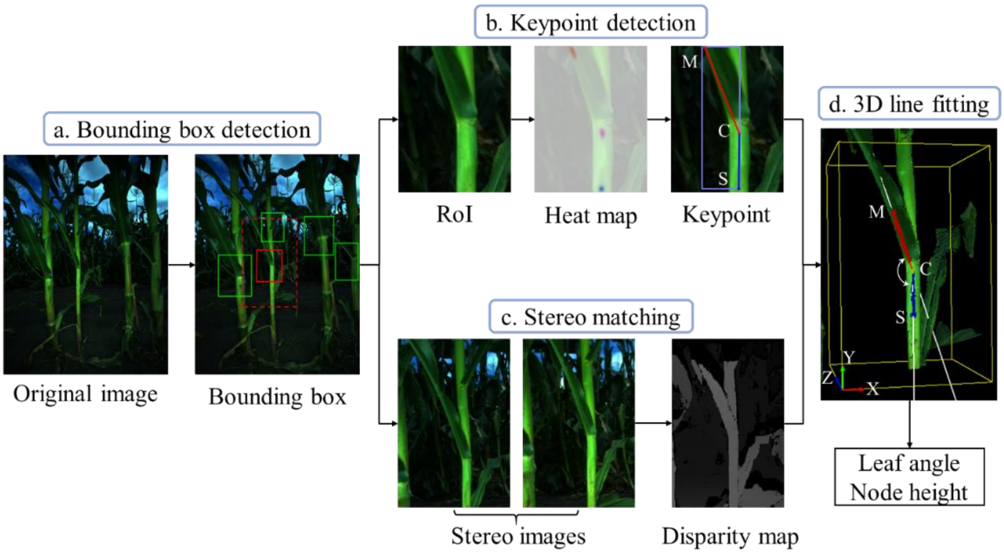


Figure 4.5. Automated image processing workflow of AngleNet for the characterization of 3D leaf angle. (a) Detection of RoIs denotated by solid bounding boxes. The red solid

bounding box is used as an example to illustrate the sequential processing steps. The red dashed bounding box is the expanded area used for stereo matching. (b) Keypoint detection. Three keypoints are extracted from the heat map of the RoI: a point near the leaf collar (C), a point on the midrib (M), and a point on the stem (S). (c) Stereo matching is carried out on the cropped stereoscopic images, and a disparity map is generated. (d) A 3D model reconstructed from the disparity map and color images. The detection results in 2D images were back-projected to the 3D space, and two lines (white) were fitted to compute the actual leaf angle. The supplementary angle of $\angle$MCS was defined as the leaf angle in this study. The pipeline outputs two types of architecture-related traits: leaf angle and its associated node height.

4.2.3.1. Keypoint Triplets for Leaf Angle Detection

An angle in 2D or 3D can be specified by a triplet of points and their topology. For leaf angle measurements, the triplet includes a point near the leaf collar (C), a point on the midrib (M), and a point on the stem (S). In this study, we aimed to identify the three keypoints with predefined relationships of a leaf angle in 2D images and reproject them into the 3D space to quantify the actual leaf angle. The region near a leaf collar, where a leaf intersects with a stem, has a unique appearance and provides a well-defined condition for the detection of keypoint C. However, defining keypoints S and M can be ambiguous because many homogeneous points along the direction of the midrib can be used to define a leaf angle. To solve this problem, we used a bounding box centered at leaf collar with a size around $0.1m \times 0.1m$ define the RoI, which enclosed the region used for in-field measurements (Figure 4.4a). Subsequently, the point on the midrib and close to the border of the bounding box was defined as M. The same principle was used for the annotation of keypoint S. In AngleNet, a RoI near each leaf collar was first detected as a bounding box (Figure 4.5a). Subsequently, the locations of keypoints M and S became well-defined with respect to the collar and the boundary, and the triplet in the region inside the object bounding box could then be identified (Figure 4.5b).

Anchor-based detectors, such as Faster R-CNN (Ren et al., 2017) and YOLOv3 (Redmon & Farhadi, 2018), enumerate a nearly exhaustive set of anchor boxes and classify each of them, which can be inefficient and computationally complex. To overcome the drawbacks of anchor-based approaches, various anchor-free detectors that model an object as points instead of using a bounding box have been developed (Kong et al., 2020; Law & Deng, 2018). For example, CenterNet (Zhou et al., 2019), which represents each instance based on its features at the center point, is a simple, fast, and accurate anchor-free detector. In the current study, we employed an anchor-free detector to detect the regions and keypoints of interest for leaf angle characterization. The results were obtained using a modified CenterNet architecture. The region near a leaf collar was modeled as a single point located at $p = (\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2})$, where (x_1, y_1) and (x_2, y_2) are the top left and bottom right points of the bounding box of the region, respectively. The detection of the object center is then transformed into a keypoint detection problem. In this study, the model takes an image of size $W \times H$ as input and aims to produce a $\frac{W}{R} \times \frac{H}{R} \times C$ heatmap. Here, R represents the output stride and is set as 4 in this study, and C is the number of classes (1 for collar detection and 3 for leaf angle keypoint detection). For the ground truth of heatmap, the values at object center are set to 1 and the other negative samples are set as 0. A single network is used to predict the keypoints $\hat{Y}$, offsets $\hat{O}$, and size $\hat{S}$. To train a network, each center point of object k is rendered by a Gaussian-shaped peak Y_{xy} and trained with focal loss L_k (Eq. 4.1). The local maximums in the heatmap are the estimated centers, and height and width are predicted based on the images at each center. In addition, a local offset loss L_{off} (Eq. 4.2) is trained to recover the discretization error caused by the resampling process. Finally, the object size regressed from the center point is trained with an L_1 loss L_{size} (Eq. 4.3). The overall loss is the weighted sum of three loss terms: focal loss (L_k), local offset loss (L_{off}), and size loss (L_{size}).The locations of the three keypoints (C, S, and M) that form the leaf angle are regressed as relative displacements from the center. The ground-truth keypoint heat maps are trained with focal loss and local offset loss via a process analogous to that used for center detection.

$$L_k = \frac{-1}{N} \sum_{xy} \begin{cases} (1-\hat{Y}_{xy})^{\alpha} \log(\hat{Y}_{xy}) & \text{if } Y_{xy} = 1 \\ (1-Y_{xy})^{\beta} (\hat{Y}_{xy})^{\alpha} \log(1-\hat{Y}_{xy}) & \text{otherwise} \end{cases} \tag{4.1}$$

$$L_{\mathrm{off}} = \frac{1}{N} \sum_{p} \left| \hat{O}_{\tilde{p}} - (\frac{p}{R} - \tilde{p}) \right| \tag{4.2}$$

$$L_{\mathrm{size}} = \frac{1}{N} \sum_{k=1}^{N} \left| \hat{S}_{p_k} - S_k \right| \tag{4.3}$$

where α and β are hyper-parameters, N is the number of keypoints, and R is the output stride size. Two backbone networks were tested for this model: deep layer aggregation (DLA-34) (Yu et al., 2017; Zhou et al., 2019) and stacked hourglass network (HG-104) (Law & Deng, 2018; Newell et al., 2016). DLA-34 is an image classification network that implements deep layer aggregation structures to better fuse feature hierarchy across layers. The aggregation architecture employs hierarchical and iterative skip connections to encompass and extend densely connected networks and feature pyramid network. HG-104 is a fully convolutional neural network that consists of two sequential hourglass modules, each containing symmetric down-sample and up-sample CNN layers with skip connections; originally used for human pose estimation (Newell et al., 2016), it has achieved state-of-the-art performance on other keypoint detection tasks (Wei et al., 2020; Zhang et al., 2018).

Two different datasets with rectified stereo images were annotated for training and testing. In the first dataset, 620 images were manually labeled with a bounding box around each leaf collar. The second dataset contained 240 images, in which each leaf angle was marked with a triplet of three dots. The annotated images were randomly split into training, validation, and test datasets at a ratio of 7:2:1. To further increase the diversity of the input images and the robustness of the network, we applied image augmentation techniques including randomly flip 50% of the images, randomly scaling (between 0.8 and 1.2), and horizontal/vertical shifting (between 0.9 and 1.1). The Adam optimizer (Kingma & Ba, 2014) was used to minimize the loss function of AngleNet. The trained models that achieved the best performance on the test dataset were used for leaf angle characterization. AngleNet is implemented with PyTorch (Paszke et al., 2017). A high-performance workstation with a NVIDIA Titan Xp GPU, a 2.2-GHz Xeon Gold 5120 CPU, and 32 GB RAM were used to train the models.

4.2.3.2. 3D Reconstruction of the Canopy

The 3D reconstruction process involves four steps: calibration, rectification, stereo matching, and triangulation. Stereo matching is particularly crucial for reconstructing a dense 3D canopy model. In the first step, each stereo camera is calibrated with a checkerboard pattern to obtain the intrinsic parameters (e.g., focal length, pixel size, principal point, and distortion coefficients) and extrinsic parameters (e.g., camera pose). The lens distortion is then removed and the stereoscopic image pair is rectified using the intrinsic parameters so that the two image planes are row-aligned. Subsequently, stereo matching takes the rectified image pair as input and computes a disparity map by finding corresponding points between the left and right images. The disparities are inversely proportional to the depth values (Figure 4.5c). Finally, the process of triangulation back-projects the disparity map to 3D coordinates to generate a point cloud (Figure 4.5d). Here, the background in the reconstructed point cloud was removed by filtering out points with depth values >0.8 m.

Stereo matching is a photogrammetric technique that reconstructs depth information based on a stereoscopic image pair (Mehltretter & Heipke, 2021). Compared with structure from motion (SfM), a commonly used 3D reconstruction technique in photogrammetry, two-view stereo imaging requires a minimum number of input images, which makes it practical for in-field applications. Many traditional algorithms, such as semi-global matching (Hirschmüller, 2008), have been developed to detect stereo correspondence. However, stereo matching for field crops can still be challenging due to low-texture surfaces, heavy occlusions, and variable lighting conditions. Recent breakthroughs in machine learning and deep learning techniques have achieved impressive results in stereo correspondence matching (Poggi et al., 2020). One of the most impactful works in this area describes MC-CNN (Zbontar & Lecun, 2016), a CNN-trained algorithm to robustly predict the similarity of two image patches. In the current study, we used a state-of-the-art stereo matching method known as 3D-MST (Li et al., 2017) to compute disparity images for the input stereo pairs and reconstruct the dense canopy. This algorithm uses MC-CNN for matching cost computation and features a cost-aggregation method with minimum spinning tree (MST)-based support region filtering. A multi-MST structure was developed to reduce the computational complexity, and a PatchMatch random search strategy was implemented to efficiently find the 3D labels of each pixel. 3D-MST is one of the top-ranking algorithms in the Middlebury 3.0 benchmark (https://vision.middlebury.edu/stereo/) and has generated convincing results for reconstructing dense plant canopies in the field (Bao et al., 2019a).

4.2.3.3. Trait Extraction

2D to 3D reprojection

Through keypoint detection, a triplet of points (C, M, and S) and two edges (CM and CS) were created for each leaf angle (Figure 4.5b). These detection results were reprojected into the 3D point cloud using the intrinsic parameters of the stereo camera (Figure 4.5d). After reprojection, the 3D positions of the targets were represented in the left camera's rectified coordinate system, where the origin is located at the projection center of the left camera and the *y* axis represents the direction perpendicular to the ground plane pointing upwards.

Estimation of leaf angle

Ideally, the 3D coordinates of the three keypoints can be used for angle measurement. However, there are some cases related to the absence of depth measurements or problematic depth measurement with large errors in some pixels. To increase the robustness of measuring an angle, the two edges (CM and CS) formed by the three keypoints were used to estimate leaf angle. Co-registration of color and 3D images provides the 3D locations of the two edges. However, projecting a line segment to a point cloud results in a dimensionality increment, an operation that does not generally result in a unique 3D line. In addition, it is possible that some remaining noise might be present in the resulting 3D patch in the point cloud. Such noise points can arise from curved leaf blades, occlusions, or stereo matching errors. To robustly determine the 3D locations of the CM and CS edges, the RANSAC algorithm (Fischler & Bolles, 1981) was implemented to fit lines in the 3D space. The algorithm iteratively fits a line based on two randomly selected points. The line fit with the most inliers is regarded as the final fit. For each edge, the RANSAC line fitting algorithm takes all the candidate points in the 3D patch (the red/blue points in Figure 4.5d) as input and outputs a position vector $\boldsymbol{p}$ and a direction vector $\boldsymbol{d}$ that determine a line.

The RANSAC-based line fitting algorithm can deal with a moderate amount of noise in the point cloud. However, we observed that when a 2D line was reprojected onto a curved leaf blade, the corresponding 3D patch tended to be discontinuous in the 3D space. In this case, the fitted line generated by the RANSAC algorithm did not effectively represent the direction of the midrib. To solve this problem, a fittings score was defined to remove these lines. The fitting score was defined as the ratio of the inliers to the total number of input points, where point $\boldsymbol{pt}$ was regarded as an inlier if its distance (Eq. 4.4) to the detected line was <0.005 m. An edge was regarded as valid for a leaf angle measurement if the fitting score was larger than 0.8; otherwise, the fit and the corresponding 3D patch were rejected from further analysis. Finally, the direction vectors of two valid edges were used to estimate leaf angle (Eq. 4.5).

$$\text{dis} = \frac{\|\boldsymbol{d} \times (\boldsymbol{p} - \boldsymbol{pt})\|}{\|\boldsymbol{d}\|} \tag{4.4}$$

$$\alpha = \frac{\arccos\left(\left|\frac{\boldsymbol{d_s} \cdot \boldsymbol{d_m}}{\|\boldsymbol{d_s}\|\|\boldsymbol{d_m}\|}\right|\right)}{\pi} * 180° \tag{4.5}$$

where $\boldsymbol{d_s}$ and $\boldsymbol{d_m}$ are the direction vectors of CM and CS in the 3D space, respectively, $\|\boldsymbol{v}\|$ is the Euclidean norm of the vector $\boldsymbol{v}$, $|x|$ is the absolute value of x, and α is the leaf angle.

Estimation of node height

Node height is defined as the distance in the vertical direction between the leaf collar and the ground plane. In this study, the *y* coordinate of the detected keypoint C was used to calculate node height. However, the corresponding *y* coordinates of point C in the 3D point cloud are based on the local coordinate system of each stereo camera. To generate comparable node height values, an offset value was added to the *y* coordinates based on the position of the camera with respect to the ground.

4.2.4. Accuracy Assessment

The performance of the newly developed algorithm was evaluated in terms of its accuracy in accomplishing two critical steps: the detection of leaf collars in 2D images, and the estimation of leaf angle and node height in 3D point clouds.

4.2.4.1. Detection Accuracy Assessment

The performance of the AngleNet model in bounding-box detection was evaluated by average precision (AP) (Padilla et al., 2021). In addition, intersection over union (IoU) was used as an indicator of position and shape accuracy. IoU is defined as the ratio of intersection area over the union area of the predicted value and ground truth.

To evaluate the performance of keypoint detection, pixel error (PE) and normalized error (NE) were utilized. Pixel error (Eq. 4.6) is defined as the Euclidean distance between the predicted point (x, y) and the ground truth $(x_{\mathrm{gt}}, y_{\mathrm{gt}})$. Normalized error (Eq. 4.7) is calculated as the PE value normalized by the length of the leaf angle skeleton.

$$\mathrm{PE} = \sqrt{(x - x_{\mathrm{gt}})^2 + (y - y_{\mathrm{gt}})^2} \tag{4.6}$$

$$\mathrm{NE} = \frac{\mathrm{PE}}{|CS| + |CM|} \times 100\% \tag{4.7}$$

where $|CS|$ and $|CM|$ are the lengths of line segments CS and CM (Figure 4.5b), respectively, in pixels.

4.2.4.2. Accuracy Assessment of Leaf Angle and Node Height Estimations

We investigated the accuracy of leaf angle measurements by comparing the AngleNet-derived values, ground truth measured in the field, and ground truth manually measured in the 3D point clouds, as described in section 2.1.2. The node height was evaluated by comparing system-derived measurements to in-field manual measurements. The correlation coefficient (r), coefficient of determination (R^2), and root mean square error (RMSE, Eq. 4.8) were used as statistical metrics to evaluate the system performance in estimating leaf angle and node height.

$$\mathrm{RMSE} = \sqrt{mean(y - y_i)^2} \tag{4.8}$$

where y and y_i denote variable values measured by AngleNet and values from manual measurements, respectively, and n is the total number of measurements.

4.3. Results

PhenoBot with multiple tiers of PhenoStereo cameras located at different heights was able to cover the entire maize canopy. Representative examples captured by the four camera sets (PT1–PT4) are shown in Figure 4.6. Overall, the AngleNet model successfully addressed various challenges in leaf angle measurement. AngleNet was robust to diverse leaf poses relative to the camera, which resulted in different viewing angles, including the (1) side view of the lower side of the leaf blade (Figure 4.6a, PT1); (2) side view of the upper side of the leaf blade (Figure 4.6a, PT2); (3) front view of the lower side of the leaf blade (Figure 4.6a, PT3); and (4) front view of the upper side of the leaf blade (Figure 4.6a, PT4). Among these viewing angles, (1) and (4) were optimal cases because the midrib was fully exposed to the camera and the leaf angle was located at a plane parallel to the image plane. PT2 posed a challenge for the detection of keypoint M due to poor midrib visibility, while PT3 required an accurate 3D model for measuring the true leaf angle because the leaf pose was perpendicular to the image plane. The results illustrate that the trained CNN model could accurately detect RoIs and keypoints for leaf angles with various leaf orientations and under different illumination and occlusion conditions.

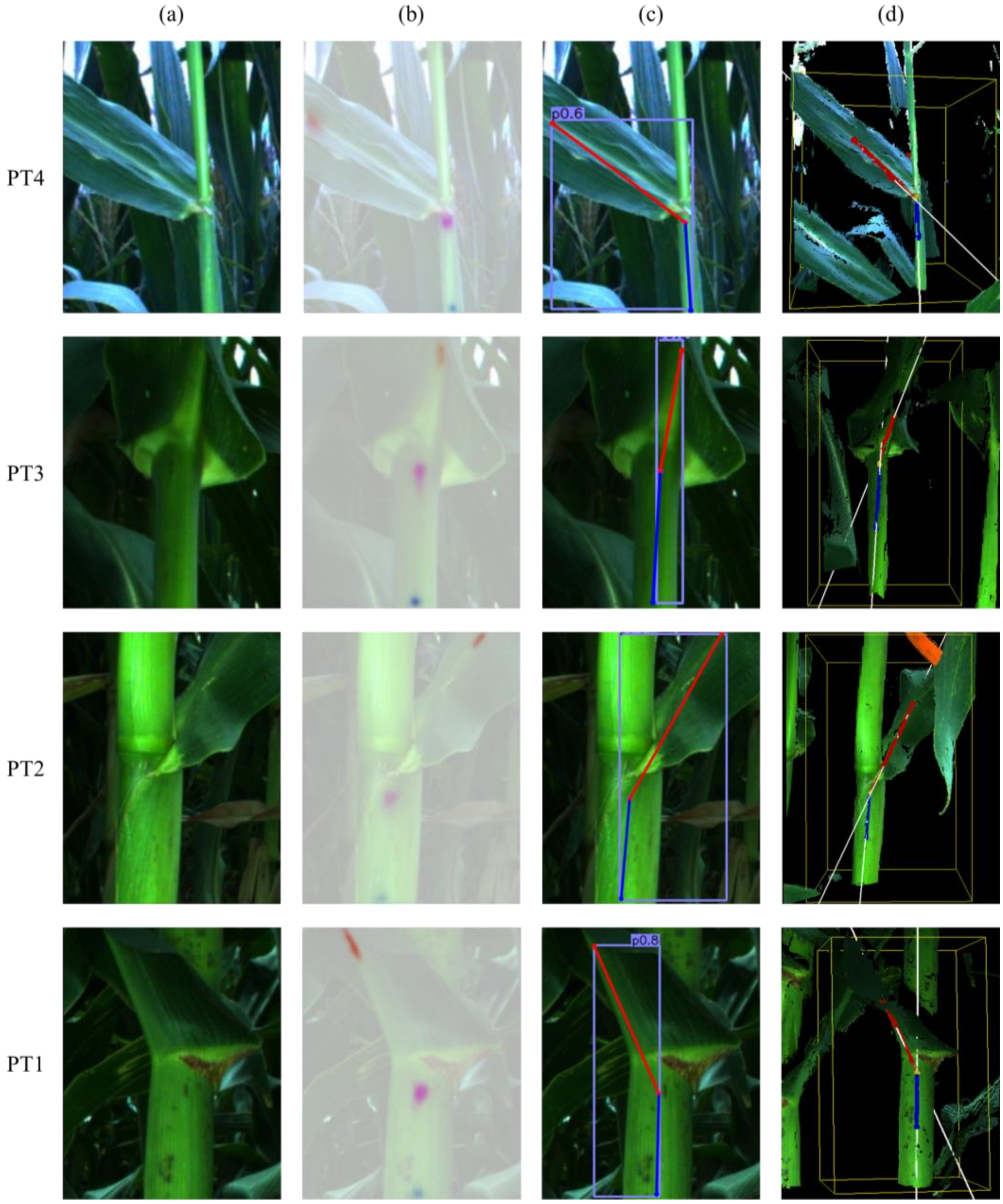


Figure 4.6. Typical examples of leaf angle measurements with intermediate results along the entire height of maize plants. (a) The RoIs detected from original images, which were

generated by the four camera heads at different heights. (b) Heat maps produced by the AngleNet model, representing the probability that each pixel represents the location of each keypoint. (c) Detection results extracted from the prediction heat map. The red and blue edges represent the directions of the midrib and stem, respectively. (d) Leaf angle measurements in 3D point cloud, where the white lines are fitted from the two detected edges in 3D space.

Figure 4.6d shows the qualitative results of 3D reconstruction and line fitting. Despite the background complexity and exposure differences, the stereo matching algorithm effectively reconstructed 3D models of the maize canopy across different heights. In addition, 3D-MST performed quite well in handling leaf blades with a homogeneous appearance (color and texture), which allowed us to reconstruct accurate 3D models with smooth and continuous plant surfaces. This superior reconstruction quality provided a reliable 3D point distribution for RANSAC line fitting. The RANSAC-based algorithm robustly detected the line that represented the direction of a midrib/stem (Figure 4.6d). In addition, the algorithm was insensitive to the random noise introduced by the reprojection of a 2D line to a 3D patch.

4.3.1. Detection Performance

We evaluated the object detection performance of the AngleNet model at two stages. During the first stage, the RoI is localized by detecting a rectangular bounding box, and during the second stage, the triplet of keypoints is detected. Two different feature extraction backbones were tested, and the one with optimal performance was used for AngleNet. Figure 4.7 shows the Precision-Recall curves of the proposed model trained with DLA-34 and HG-104. The model trained with DLA-34 can produce more accurate predictions than HG-104, supported by a higher precision value at the same recall value and IoU threshold. However, the model trained with HG-104 can achieve higher maximum recall values, indicating the model has a higher ability to find all the relevant cases. Overall, the model trained with HG-104 backbone produced better results, as illustrated by the larger AP values both at $\text{IoU}_{(0.5)}$ and $\text{IoU}_{(0.7)}$. The trained model achieved an AP of 0.908 at $\text{IoU}_{(0.5)}$ on the test dataset, highlighting the effectiveness of the model in detecting the RoI for leaf angle measurements.

For the second stage, the detector with an HG-104 backbone successfully identified 279 of 282 angles, while the model trained with the DLA-34 backbone detected 263 samples. No false detections were produced by either model. Based on the detection rate, the HG-104 was used as the backbone model for keypoint detection in AngleNet.

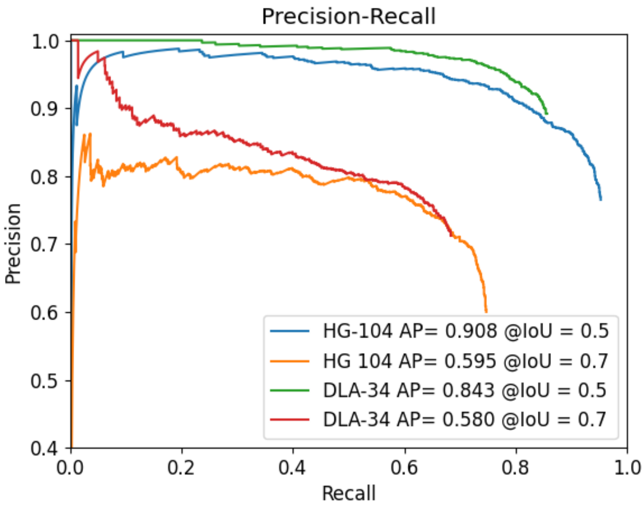


Figure 4.7. Precision-Recall (PR) curve of models trained with DLA-34 and HG-104 in detecting leaf collars on the test dataset with different IoU thresholds and corresponding AP values. An ideal PR curve should pass through the top-right corner (i.e., 100% for both precision and recall).

To evaluate the localization accuracy, we calculated the mean and standard deviation (Table 4.2) for PE and NE described in section 2.3.1. Overall, the deviations were less than 3% of the length of the leaf angle skeleton for all keypoints, indicating strong model robustness. Keypoint M produced higher PE and NE than the two other keypoints. The largest pixel errors for the three classes of keypoints in the test dataset were mainly caused by occlusions (Figure 4.8). However, as shown in the examples in Figures 4.6 and 4.8, most deviations were along the direction of the midrib/stem, which has minimal impact on angle measurement. To statistically evaluate the effectiveness of the keypoint predictions, we compared the angles estimated based on the detection results to the angles computed based on manually labeled keypoints. The mean and standard deviation of the angle errors were 1.94° and 1.67°, respectively.

Table 4.2. Euclidean distance (in pixels) between the detected keypoints and ground truth normalized by the length of the leaf angle skeleton.

	PE		NE (%)	
Keypoint	Mean	Standard deviation	Mean	Standard deviation
C	6.55	5.18	1.78	1.29
M	10.41	10.08	2.93	3.19
S	5.52	4.45	1.52	1.26

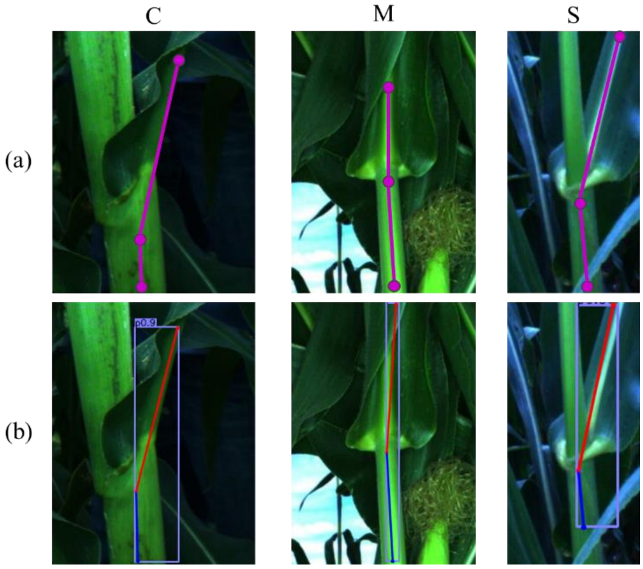
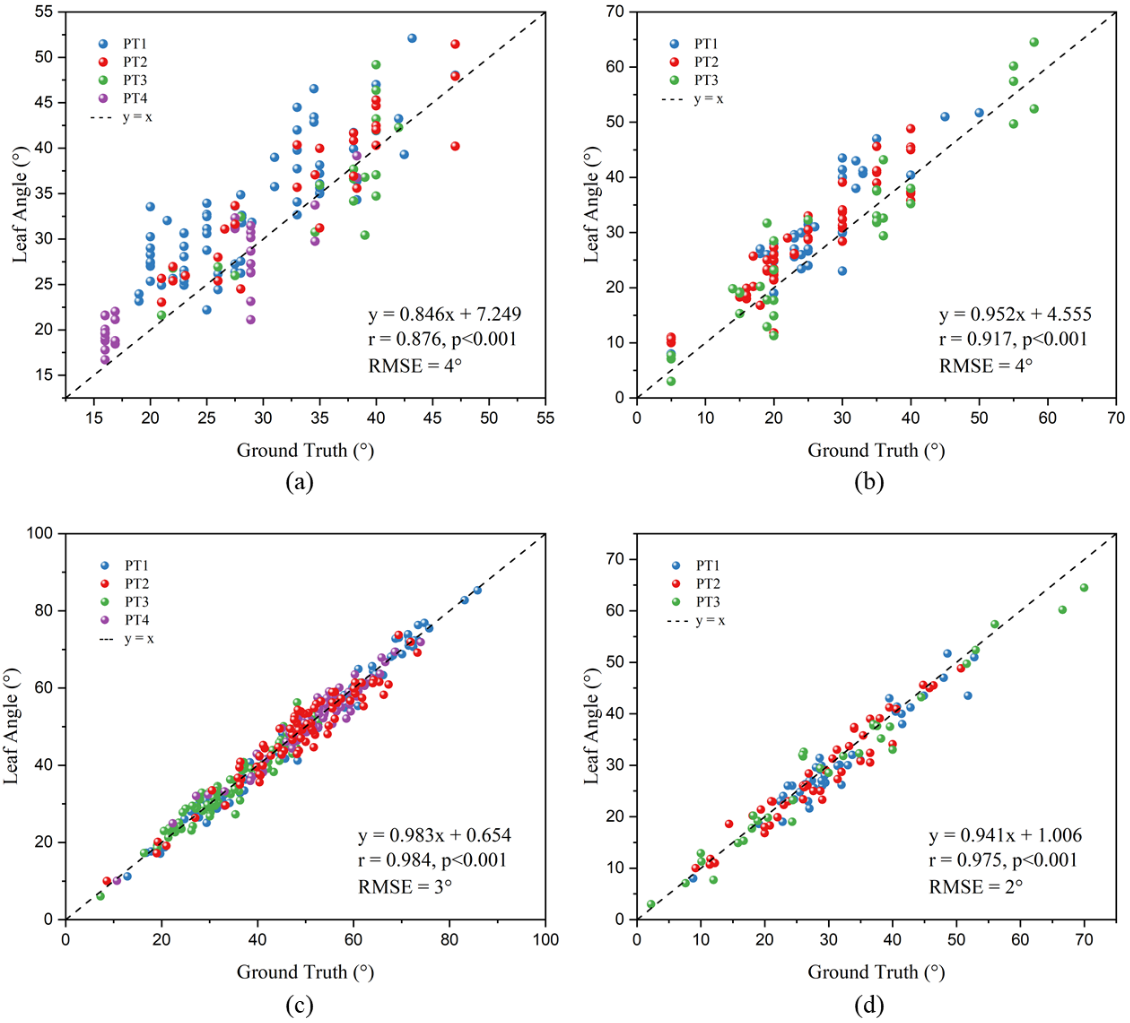


Figure 4.8. The largest pixel errors of keypoints C, M, and S in the test dataset. (a) Manually labeled ground truth. (b) Keypoint detection results produced by AngleNet.

4.3.2. Accuracy of Leaf Angle Estimates

The leaf angle estimates obtained using the current approach were highly correlated ($r > 0.87$, RMSE ≤ 4°) with in-field measurements (Figures 4.9a and 4.9b). For both datasets, the intercept of the fit line was positive, indicating a general overestimation of leaf angles compared to the ground truth. The errors in leaf angle estimation were due to a number of factors, including keypoint location errors, inaccurate line fitting, and the inability to reconstruct an accurate 3D model. Another possible source of random error is errors in ground-truth collection. Compared with field-based ground truth, the estimated leaf angles had a stronger correlation with the references from the point cloud data (Figures 4.9c and 4.9d). Possible reasons for this include the following: (1) the second ground-truth collected from point cloud excludes the errors caused by 3D reconstruction, and (2) due to the time interval between imaging and measuring, the leaf angle may have changed physically due to differences in wind and sunlight conditions, thereby affecting in-field ground-truth measurements.

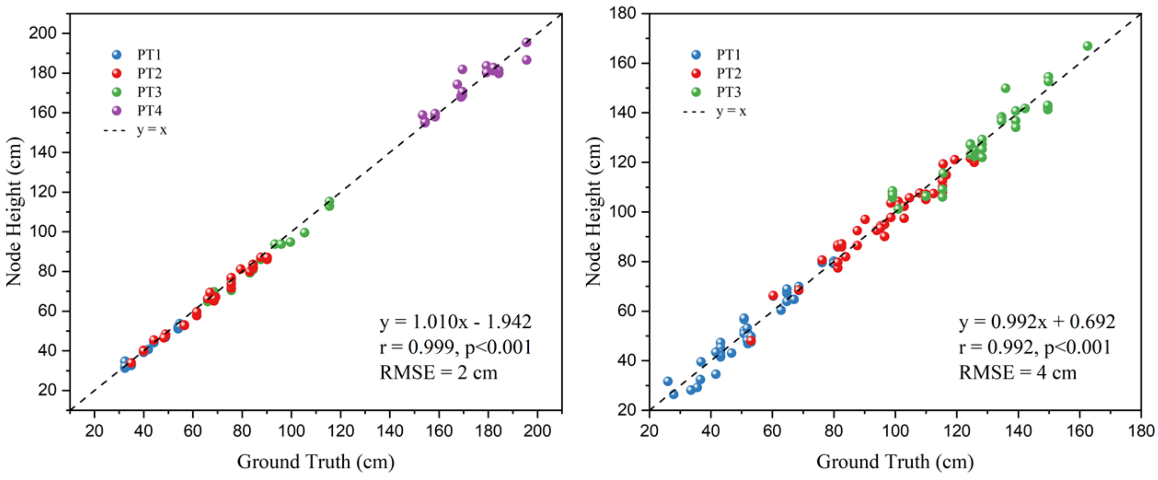


(a) (b) (c) (d)

Figure 4.9. Comparison of AngleNet-derived leaf angles with reference measurements. (a-b) Comparison of estimated leaf angles with in-field manual measurements collected using a protractor of PS4-Boone and OS3-Ames, respectively. (c-d) Comparison of estimated leaf angles with reference measurements manually measured from the 3D point clouds of PS4-Boone and PS3-Ames, respectively.

4.3.3. Accuracy of Node Height Values

Overall, the node heights calculated using the current approach were highly correlated ($r > 0.99$) with the manual measurements made in the field (Figure 4.10). The slope of the regression equation was close to 1, suggesting that system-derived node height values can be used as a direct estimate of the locations of leaf angles. We performed statistical analysis of the data collected by different camera heads (Tables 4.3 and 4.4), finding that the highest camera (PT4 in PS4-Boone and PT3 in PS3-Ames) had a significantly lower correlation coefficient than the other lower camera sets. The main reason for this difference is that the leaf collars imaged by the higher camera were located at relatively high positions, making it difficult to manually measure the node height in an accurate manner. Additionally, the top leaves of maize plants are less rigid than the bottom leaves, making them more sensitive to the movements caused by wind and PhenoBot. In general, all cameras in two datasets achieved satisfactory performance, with a high correlation ($r > 0.94$) and a low root mean square error (RMSE ≤ 4 cm). However, the model has a better performance in PS4-Boone than in PS3-Ames. Compared to PS4-Boone, the data collection in PS3-Ames has faster driving speed and longer camera-to-object distance, therefore produced stronger vibrations of the sensor mast and larger accumulated errors.



(a) (b)

Figure 4.10. Comparison of AngleNet-derived node height values with manual measurements for PS4-Boone and PS3-Ames, respectively.

Table 4.3. Comparison of the node height estimation accuracies of different camera heads for PS4-Boone.

Camera	Linear Fitting	*r*	RMSE (cm)
PT1	$y = 0.928x + 2.243$	0.992	1
PT2	$y = 0.974x + 0.306$	0.994	2
PT3	$y = 0.963x + 1.041$	0.991	2
PT4	$y = 0.867x + 23.317$	0.956	4

Table 4.4. Comparison of the node height estimation accuracies of different camera heads for PS3-Ames.

Camera	Linear Fitting	r	RMSE (cm)
PT1	$y = 1.014x - 0.830$	0.982	4
PT2	$y = 0.938x + 6.009$	0.971	4
PT3	$y = 0.948x + 6.641$	0.949	5

4.3.4. Leaf Angle Distribution in B73 and Mo17

Finally, we used the AngleNet pipeline to explore the variations in leaf angle between two inbred maize lines: B73 and Mo17. To analyze the distribution of leaf angles throughout the canopy, we performed second-order polynomial fitting between leaf angle and the corresponding node height for each line (Figure 4.11). B73 exhibited erect leaves in the upper canopy and horizontal leaf angles in the lower canopy, which is consistent with a previous report (Dzievit et al., 2019). More specifically, the leaf angle of B73 had a negative relationship with node height. This configuration, with increasingly upright leaves from the bottom to the top, maximizes the potential for light capture and increases the photosynthetic conversion efficiency (Zhang et al., 2017). Mo17 showed horizontal leaf angles near the top and bottom of the canopy but relatively erect architecture in the middle part of the canopy, especially around ears. As shown in the sample images captured by PT4 (Figure 4.12), the uppermost leaf angles of Mo17 were obviously larger than those of B73. Several angles were successfully detected in 2D images but failed to be measured in 3D space, primarily as a result of one of two situations: (1) the 2D edge was reprojected onto a wavy-shaped leaf blade and was rejected in 3D space because of a low fitting score; or (2) the angle was located near the boundary of one image and was absent from the other image of a stereo pair.

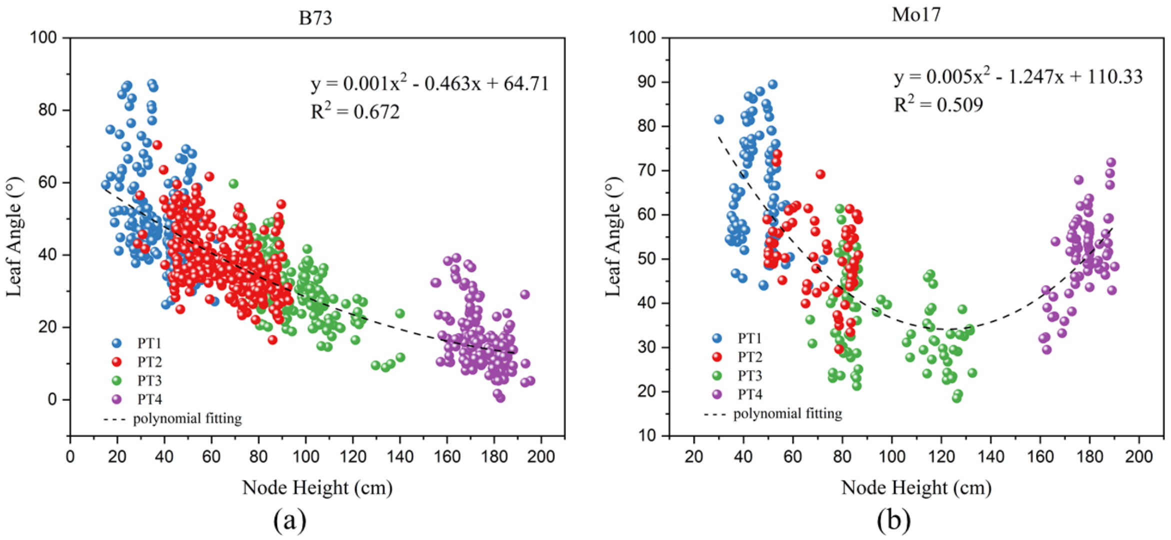


Figure 4.11. Leaf angle distribution along the entire height of B73 and Mo17 based on data generated by AngleNet. (a) The leaf angle of B73 decreased with increasing height. (b) The leaf angle of Mo17 initially decreased in the lower canopy and then increased in the upper canopy.

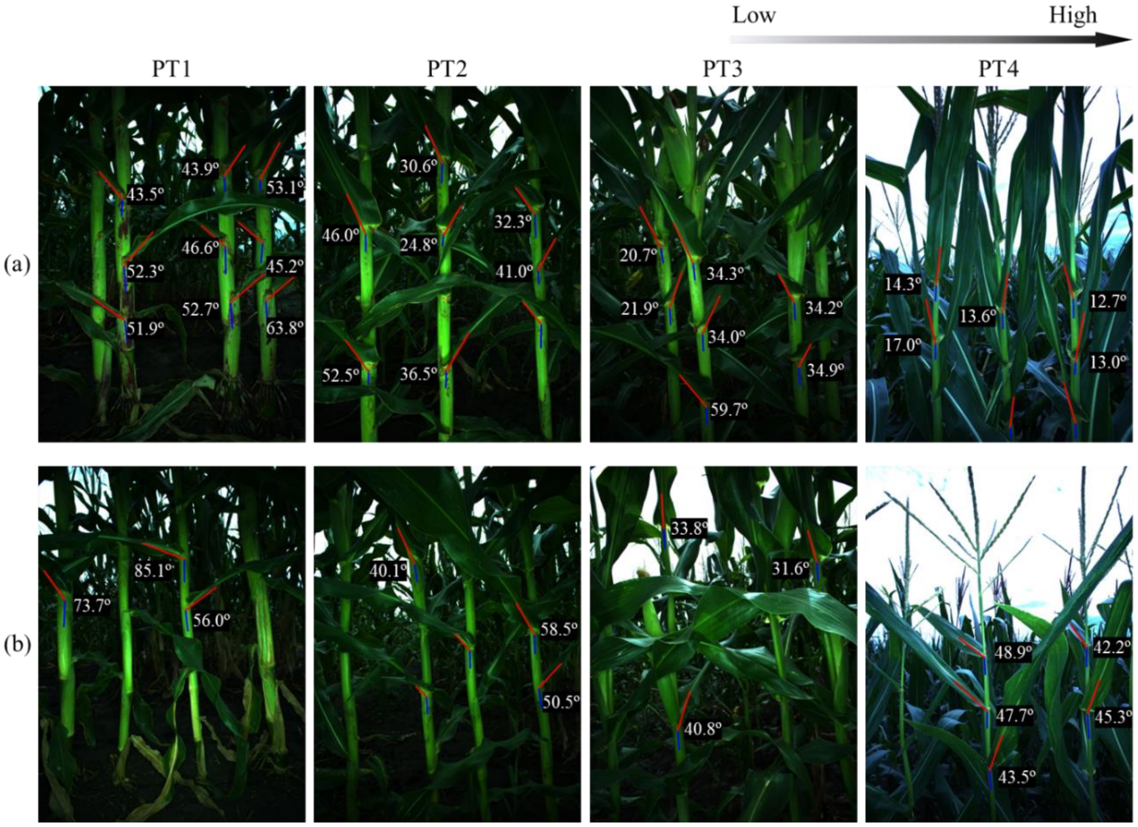


Figure 4.12. Sample outputs generated by the AngleNet pipeline showing comparisons of leaf angle at different heights between B73 and Mo17. The red and blue lines represent the directions of the midrib and stem, respectively; their supplementary angle was measured as the leaf angle. (a) B73 exhibited gradually decreasing angles starting from the bottom and had upright leaves at the top. (b) Mo17 had more horizontal leaf architecture throughout the canopy, with relatively erect architecture around ears.

4.4. Discussion

In the current study, the use of PhenoBot equipped with multiple tiers of PhenoStereo cameras provided an efficient and effective solution for high-resolution RGB and depth imaging of tall-growing crops in the field. However, expanded leaves occasionally blocked the view of the middle cameras due to narrow spacing. One possible solution to this problem is to increase the number of stereo cameras vertically. In this way, a leaf angle may be invisible to one camera but visible to another camera at a different height. Another alternative involves placing two camera sets with varying horizontal viewing angles at the same height, which could potentially mitigate the obstruction of camera views and the occlusions resulting from different plant orientations and leaf poses.

Most sophisticated approaches used to extract leaf angles from 2D and 3D images involve segmenting and skeletonizing maize plants to analyze plant architecture (Cabrera-Bosquet et al., 2016; Das Choudhury et al., 2018; Wu et al., 2019, 2020; Zermas et al., 2020). Those methods reply on visible leaf tips for leaf segmentation (Gaillard et al., 2020; Souza et al., 2021) and need to adjust a set of parameters to achieve satisfactory performance (Wu et al., 2019; Zermas et al., 2020; Zhu et al., 202), which are not feasible for large-scale field-based applications due to the space resolution of the raw data and occlusion of nearby plants. The AngleNet pipeline proposed in this study focuses on the regions near leaf collars, which bypasses the complex procedures of individual plant and leaf segmentation. The leaf angle is located based on its associated height. Inspired by human pose estimation, this is the first study to implement the topology of keypoints for leaf angle measurement. The AngleNet pipeline showed the ability to quantify leaf angle in 3D space regardless of leaf orientation relative to the camera, along with robustness to complex and varying outdoor conditions. By combining the PhenoBot platform with the AngleNet pipeline, the distribution of leaf angles across the canopy could be investigated nondestructively under natural conditions at multiple time points. Furthermore, this framework could easily be modified to measure the angles of other objects, such as sorghum plants and even trees in forestry studies. Beyond angle measurements, the concept of keypoint detection could be extended to length measurements, such as plant height and stem length.

Compared to traditional skeletonization methods, the major disadvantage of AngleNet is that it cannot deal with wavy-shaped leaves (Figure 4.13a). Since we focused on a small local area, the presence of a partially rolling leaf blade would result in a highly curved patch during the process of reprojecting a 2D line to a 3D point cloud, making it difficult and unreliable to fit the direction line in the resulting point cloud. This limitation could be overcome by performing keypoint detection on both left and right images and reconstructing the 3D line based on a pair of 2D line segments. Additionally, we observed that the 2D line was not exactly parallel to the direction of the midrib/stem (Figure 4.13b, left), which resulted in an inversible error for angle measurements in 3D space. Therefore, the keypoint locations and line directions should be refined before reprojecting to 3D coordinates. For example, locally straight contours (Lee et al., 2013) could be detected and used as a reference to adjust the line directions (Figure 4.13b, right).

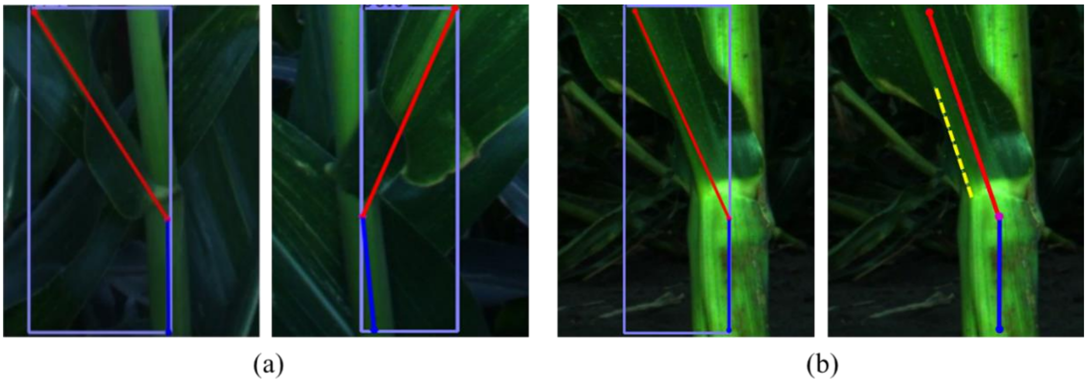


Figure 4.13. Limitations of AngleNet. (a) The midrib direction lines were located on a rolling leaf, where the resulting point cloud in 3D was not suitable for line fitting. (b) An alternative way to refine leaf angle topology in a 2D image. Left: the initial detection results produced by AngleNet. The red line is not exactly parallel to the direction of the midrib. Right: the direction of the midrib was refined based on a local edge (yellow dashed line).

4.5. Conclusion

In this study, we employed an automated system to characterize leaf angle in agronomically grown maize plants in the field. A novel leaf angle detection and measurement pipeline, named AngleNet, was developed to obtain actual quantitative data for leaf angle throughout the entire height of the plant canopy. This pipeline utilizes deep CNNs to detect leaf angle as a triplet of keypoints and stereo matching algorithms to quantify leaf angle in the reconstructed 3D point cloud. Based on the detection and reconstruction results, the AngleNet-derived leaf angle and the associated node height were highly correlated with ground truth. In addition, the framework was successfully implemented to explore the variation of leaf angle in shoot architecture of two maize inbred lines, B73 and Mo17. The results of quantitative analysis were consistent with the actual leaf angle distributions along plant height in these lines. Specifically, B73 has more erect architecture in the upper top canopy than Mo17. Stereo vision with the proper customization represents a practical tool to rapidly acquire high-quality RGB images and measure plant morphology under field conditions. The proposed system represents a feasible way to automatically quantify leaf angle in maize plants in a nondestructive manner. In addition, this system provides a new strategy for breeders to optimize plant architecture toward a smart canopy in cereal crops through large-scale field experiments.

Acknowledgements

This project was funded by the U.S. National Science Foundation Major Research Instrumentation Program (Award No: 1625364) and the Plant Sciences Institute at Iowa State University. This research is also a product of the Iowa Agriculture Experiment Station, Ames, Iowa, Project No. IOW03712, which is partially supported by Hatch Act and State of Iowa funds. We thank Taylor Tuel for his contributions to the design of PhenoBot and Gregory Schoenbaum for his assistance with field experiments.

References

Andújar, D., Dorado, J., Fernández-Quintanilla, C., & Ribeiro, A. (2016). An Approach to the Use of Depth Cameras for Weed Volume Estimation. *Sensors*, *16*(7), 972. https://doi.org/10.3390/s16070972

Apelt, F., Breuer, D., Nikoloski, Z., Stitt, M., & Kragler, F. (2015). *Phytotyping 4D : a light-field imaging system for non-invasive and accurate monitoring of spatio-temporal plant growth*. 693–706. https://doi.org/10.1111/tpj.12833

Atefi, A., Ge, Y., Pitla, S., & Schnable, J. (2019). In vivo human-like robotic phenotyping of leaf traits in maize and sorghum in greenhouse. *Computers and Electronics in Agriculture,* 163, 104854. https://doi.org/10.1016/j.compag.2019.104854

Bao, Y., Tang, L., Breitzman, M. W., Salas Fernandez, M. G., & Schnable, P. S. (2019a). 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. (2019b). 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

Cabrera-Bosquet, L., Fournier, C., Brichet, N., Welcker, C., Suard, B., & Tardieu, F. (2016). High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. *The New Phytologist*, *212*(1), 269–281. https://doi.org/10.1111/nph.14027

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

Che, Y., Wang, Q., Xie, Z., Zhou, L., Li, S., Hui, F., Wang, X., Li, B., & Ma, Y. (2020). Estimation of maize plant height and leaf area index dynamics using an unmanned aerial vehicle with oblique and nadir photography. *Annals of Botany*, *126*, 765–773. https://doi.org/10.1093/aob/mcaa097

Crick, C., Jay, G., Osentoski, S., Pitzer, B., & Jenkins, O. C. (2017). Rosbridge: ROS for non-ROS users. *Springer Tracts in Advanced Robotics*, *100*, 493–504. https://doi.org/10.1007/978-3-319-29363-9_28

Das Choudhury, S., Bashyam, S., Qiu, Y., Samal, A., & Awada, T. (2018). Holistic and component plant phenotyping using temporal image sequence. *Plant Methods*, *14*(1), 1–21. https://doi.org/10.1186/s13007-018-0303-x

Duan, T., Chapman, S. C., Holland, E., Rebetzke, G. J., Guo, Y., & Zheng, B. (2016). Dynamic quantification of canopy structure to characterize early plant vigour in wheat genotypes. *Journal of Experimental Botany*, *67*(15), 4523–4534. https://doi.org/10.1093/jxb/erw227

Dzievit, M. J., Li, X., & Yu, J. (2019). *Dissection of Leaf Angle Variation in Maize through Genetic Mapping and Meta-Analysis*. 1–12. https://doi.org/10.3835/plantgenome2018.05.0024

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision, 88*(2), 303–338. https://doi.org/10.1007/s11263-009-0275-4

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM, 24*(6), 381–395. https://doi.org/10.1145/358669.358692

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. (2020). *Navigation control of a robotic vehicle for field-based phenotyping* (Doctoral dissertation, Iowa State University).

Gai, J., Xiang, L., & Tang, L. (2021). Using a depth camera for crop row detection and mapping for under-canopy navigation of agricultural robotic vehicle. *Computers and Electronics in Agriculture, 188,* 106301. https://doi.org/10.1016/j.compag.2021.106301

Gaillard, M., Miao, C., Schnable, J., & Benes, B. (2020). Sorghum segmentation by skeleton extraction. In *European Conference on Computer Vision* (pp. 296-311). Springer, Cham. https://doi.org/10.1007/978-3-030-65414-6_21

Gélard, W., Devy, M., Herbulot, A., & Burger, P. (2017). Model-based segmentation of 3D point clouds for phenotyping sunflower plants. In *12. International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications.*

Gibbs, J. A., Pound, M., French, A. P., Wells, D. M., Murchie, E., & Pridmore, T. (2018). Plant phenotyping: an active vision cell for three-dimensional plant shoot reconstruction. *Plant physiology, 178*(2), 524-534. https://doi.org/10.1104/pp.18.00664

Guo, W., Fukano, Y., Noshita, K., & Ninomiya, S. (2020). Field-based individual plant phenotyping of herbaceous species by unmanned aerial vehicle. *Ecology and Evolution, 10*(21), 12318–12326. https://doi.org/10.1002/ece3.6861

Hammer, G. L., Dong, Z., McLean, G., Doherty, A., Messina, C., Schussler, J., Zinselmeier, C., Paszkiewicz, S., & Cooper, M. (2009). Can Changes in Canopy and/or Root System Architecture Explain Historical Maize Yield Trends in the U.S. Corn Belt? *Crop Science, 49*(1), 299–312. https://doi.org/10.2135/cropsci2008.03.0152

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

Hui, F., Zhu, J., Hu, P., Meng, L., Zhu, B., Guo, Y., ... & Ma, Y. (2018). Image-based dynamic quantification and high-accuracy 3D evaluation of canopy structure of plant populations. A*nnals of botany*, *121*(5), 1079-1088. https://doi.org/10.1093/aob/mcy016

Jiang, Y., & Li, C. (2020). Convolutional Neural Networks for Image-Based High-Throughput Plant Phenotyping: A Review. *Plant Phenomics*, *2020*, 1–22. https://doi.org/10.34133/2020/4152816

Kenchanmane Raju, S. K., Adkins, M., Enersen, A., Santana de Carvalho, D., Studer, A. J., Ganapathysubramanian, B., Schnable, P. S., & Schnable, J. C. (2020). Leaf Angle eXtractor: A high-throughput image processing framework for leaf angle measurements in maize and sorghum. *Applications in Plant Sciences*, *8*(8). https://doi.org/10.1002/aps3.11385

Kicherer, A., Herzog, K., Bendel, N., Klück, H. C., Backhaus, A., Wieland, M., ... & Töpfer, R. (2017). Phenoliner: a new field phenotyping platform for grapevine research. *Sensors*, *17*(7), 1625. https://doi.org/10.3390/s17071625

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C. (2019). Deep learning – Method overview and review of use for fruit detection and yield estimation. *Computers and Electronics in Agriculture*, *162*(April), 219–234. https://doi.org/10.1016/j.compag.2019.04.017

Kong, T., Sun, F., Liu, H., Jiang, Y., Li, L., & Shi, J. (2020). Foveabox: Beyound anchor-based object detection. *IEEE Transactions on Image Processing*, *29*, 7389-7398. https://doi.org/10.1109/TIP.2020.3002345

Law, H., & Deng, J. (2018). CornerNet: Detecting Objects as Paired Keypoints. *International Journal of Computer Vision*, *128*(3), 642–656. http://arxiv.org/abs/1808.01244

Lee, J. H., Zhang, G., Lim, J., & Suh, I. H. (2013). Place recognition using straight lines for vision-based SLAM. *Proceedings - IEEE International Conference on Robotics and Automation*, 3799–3806. https://doi.org/10.1109/ICRA.2013.6631111

Lewis, M. W., Bolduc, N., Hake, K., Htike, Y., Hay, A., Candela, H., & Hake, S. (2014). Gene regulatory interactions at lateral organ boundaries in maize. *Development (Cambridge)*, *141*(23), 4590–4597. https://doi.org/10.1242/dev.111955

Li, D., Cao, Y., Shi, G., Cai, X., Chen, Y., Wang, S., & Yan, S. (2019). An overlapping-free leaf segmentation method for plant point clouds. *IEEE Access*, *7*, 129054-129070. https://doi.org/10.1109/ACCESS.2019.2940385

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

Lu, H., Cao, Z., Xiao, Y., Fang, Z., & Zhu, Y. (2017). Towards fine-grained maize tassel flowering status recognition : Dataset , theory and practice. *Applied Soft Computing Journal*, *56,* 34–45. https://doi.org/10.1016/j.asoc.2017.02.026

Mansfield, B. D., & Mumm, R. H. (2014). Survey of Plant Density Tolerance in U.S. Maize Germplasm. *Crop Science*, *54*(1), 157–173. https://doi.org/10.2135/cropsci2013.04.0252

Mantilla-Perez, M. B., Bao, Y., Tang, L., Schnable, P. S., & Salas-Fernandez, M. G. (2020). Towards “smart canopy” sorghum: discovery of the genetic control of leaf angle across layers. *Plant Physiology,* pp.00632.2020. https://doi.org/10.1104/pp.20.00632

Mccormick, R. F., Truong, S. K., & Mullet, J. E. (2016). 3D sorghum reconstructions from depth images identify QTL regulating shoot architecture. *Plant Physiology, 172*(October), pp.00948.2016. https://doi.org/10.1104/pp.16.00948

Mehltretter, M., & Heipke, C. (2021). ISPRS Journal of Photogrammetry and Remote Sensing Aleatoric uncertainty estimation for dense stereo matching via CNN-based cost volume analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, *171*(August 2020), 63–75. https://doi.org/10.1016/j.isprsjprs.2020.11.003

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

Newell, A., Yang, K., & Deng, J. (2016). Stacked hourglass networks for human pose estimation. *European Conference on Computer Vision , 9912 LNCS*, 483–499. https://doi.org/10.1007/978-3-319-46484-8_29

Ort, D. R., Merchant, S. S., Alric, J., Barkan, A., Blankenship, R. E., Bock, R., … &Zhu, X. G. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 112, Issue 28, pp. 8529–8536). National Academy of Sciences. https://doi.org/10.1073/pnas.1424031112

Padilla, R., Passos, W. L., Dias, T. L., Netto, S. L., & da Silva, E. A. (2021). A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics, 10*(3), 279. https://doi.org/10.3390/electronics10030279

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., Facebook, Z. D., Research, A. I., Lin, Z., Desmaison, A., Antiga, L., Srl, O., & Lerer, A. (2017). *Automatic differentiation in PyTorch*.

Pérez-borrero, I., Marín-santos, D., Gegúndez-arias, M. E., & Cortés-ancos, E. (2020). A fast and accurate deep learning method for strawberry instance segmentation. *Computers and Electronics in Agriculture*, *178*(August), 105736. https://doi.org/10.1016/j.compag.2020.105736

Poggi, M., Tosi, F., Batsos, K., Mordohai, P., & Mattoccia, S. (2020). On the Synergies between Machine Learning and Stereo: a Survey. *ArXiv*. http://arxiv.org/abs/2004.08566

Redmon, J., & Farhadi, A. (2018). *YOLOv3: An Incremental Improvement*. http://arxiv.org/abs/1804.02767

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

Russell, W. A. (1972). Registration of B70 and B73 Parental Lines of Maize 1 (Reg. Nos. PL16 and PL17) . *Crop Science*, *12*(5), 721–721. https://doi.org/10.2135/cropsci1972.0011183x001200050085x

Santos, T. T., de Souza, L. L., dos Santos, A. A., & Avila, S. (2020). Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *Computers and Electronics in Agriculture*, *170*(February), 105247. https://doi.org/10.1016/j.compag.2020.105247

Stanford Artificial Intelligence Laboratory et al. (2018). Robotic Operating System. Retrieved March 20, 2021, from https://www.ros.org/

Souza, A., & Yang, Y. (2021). High-Throughput Corn Image Segmentation and Trait Extraction Using Chlorophyll Fluorescence Images. *Plant Phenomics*, 2021. https://doi.org/10.34133/2021/9792582

Sun, S., Li, C., Chee, P. W., Paterson, A. H., Jiang, Y., Xu, R., Robertson, J. S., Adhikari, J., & Shehzad, T. (2020). 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

Tang, D., Chen, Z., Ni, J., Jiang, Q., Li, P., Wang, L., Zhou, J., Li, C., & Liu, J. (2018). Identification of QTL for leaf angle at canopy-wide levels in maize. In *bioRxiv* (p. 499665). bioRxiv. https://doi.org/10.1101/499665

Thapa, S., Zhu, F., Walia, H., Yu, H., & Ge, Y. (2018). *A Novel LiDAR-Based Instrument for High-Throughput, 3D Measurement of Morphological Traits in Maize and Sorghum*. https://doi.org/10.3390/s18041187

Toris, R., Kammerl, J., Lu, D. V., Lee, J., Jenkins, O. C., Osentoski, S., Wills, M., & Chernova, S. (2015). Robot Web Tools: Efficient messaging for cloud robotics. *IEEE International Conference on Intelligent Robots and Systems*, *2015-December*, 4530–4537. https://doi.org/10.1109/IROS.2015.7354021

Truong, S. K., Mccormick, R. F., Rooney, W. L., & Mullet, J. E. (2015). *Harnessing Genetic Variation in Leaf Angle to Increase Productivity of Sorghum bicolor*. *201*(November), 1229–1238. https://doi.org/10.1534/genetics.115.178608

Tuel, T. L. (2019). *A robotic proximal sensing platform for in-field high-throughput crop phenotyping* (Master's thesis, Iowa State University).

Vit, A., Shani, G., & Bar-hillel, A. (2020). Length phenotyping with interest point detection. *Computers and Electronics in Agriculture*, *176*(July), 105629. https://doi.org/10.1016/j.compag.2020.105629

Wang, N., Cao, D., Gong, F., Ku, L., Chen, Y., & Wang, W. (2015). Differences in properties and proteomes of the midribs contribute to the size of the leaf angle in two near-isogenic maize lines. *Journal of Proteomics*, *128*, 113–122. https://doi.org/10.1016/j.jprot.2015.07.027

Wang, Y., Wen, W., Wu, S., Wang, C., Yu, Z., Guo, X., & Zhao, C. (2019). Maize plant phenotyping: Comparing 3D laser scanning, multi-view stereo reconstruction, and 3D digitizing estimates. *Remote Sensing*, *11*(1). https://doi.org/10.3390/rs11010063

Wei, H., Zhang, Y., Chang, Z., Li, H., Wang, H., & Sun, X. (2020). Oriented objects as pairs of middle lines. *ISPRS Journal of Photogrammetry and Remote Sensing*, *169*(July), 268–279. https://doi.org/10.1016/j.isprsjprs.2020.09.022

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

Wu, S., Wen, W., Xiao, B., Guo, X., Du, J., Wang, C., & Wang, Y. (2019). An Accurate Skeleton Extraction Approach From 3D Point Clouds of Maize Plants. *Frontiers in Plant Science*, *10*, 248. https://doi.org/10.3389/fpls.2019.00248

Xiang, L., Tang, L., Gai, J., & Wang, L. (2020). PhenoStereo: A high-throughput stereo vision system for field-based plant phenotyping - With an application in sorghum stem diameter estimation. *ASABE 2020 Annual International Meeting*. https://doi.org/10.13031/aim.202001190

Xiang, L., Bao, Y., Tang, L., Ortiz, D., & Salas-Fernandez, M. G. (2019). Automated morphological traits extraction for sorghum plants via 3D point cloud data analysis. Computers and Electronics in Agriculture, 162, 951-961. https://doi.org/10.1016/j.compag.2019.05.043

Xiao, S., Chai, H., Shao, K., Shen, M., Wang, Q., Wang, R., ... & Ma, Y. (2020). Image-based dynamic quantification of aboveground structure of sugar beet in field. *Remote Sensing*, *12*(2), 269. https://doi.org/10.3390/rs12020269

Yu, F., Wang, D., Shelhamer, E., & Darrell, T. (2017). Deep Layer Aggregation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2403–2412. http://arxiv.org/abs/1707.06484

Zabawa, L., Kicherer, A., Klingbeil, L., Töpfer, R., Kuhlmann, H., & Roscher, R. (2020). Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *164*, 73-83.

Zbontar, J., & Lecun, Y. (2016). Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches. In *Journal of Machine Learning Research* (Vol. 17). https://doi.org/10.5555/2946645.2946710

Zermas, D., Morellas, V., Mulla, D., & Papanikolopoulos, N. (2020). 3D model processing for high throughput phenotype extraction – the case of corn. *Computers and Electronics in Agriculture*, *172*(June 2019), 105047. https://doi.org/10.1016/j.compag.2019.105047

Zhang, X., Huang, C., Wu, D., Qiao, F., Li, W., Duan, L., Wang, K., Xiao, Y., Chen, G., Liu, Q., Xiong, L., Yang, W., & Yan, J. (2017). High-throughput phenotyping and QTL mapping reveals the genetic architecture of maize plant growth. *Plant Physiology*, *173*(3), 1554–1564. https://doi.org/10.1104/pp.16.01516

Zhang, Y., Sun, H., Zuo, J., Wang, H., Xu, G., & Sun, X. (2018). Aircraft type recognition in remote sensing images based on feature learning with conditional generative adversarial networks. *Remote Sensing*, *10*(7), 1–19. https://doi.org/10.3390/rs10071123

Zhao, X., Fang, P., Zhang, J., & Peng, Y. (2018). QTL mapping for six ear leaf architecture traits under water-stressed and well-watered conditions in maize (*Zea mays* L.). *Plant Breeding*, *137*(1), 60–72. https://doi.org/10.1111/pbr.12559

Zhou, X., Wang, D., & Krähenbühl, P. (2019). Objects as Points. *ArXiv*. http://arxiv.org/abs/1904.07850

Zhu, C., Miao, T., Xu, T., Yang, T., & Li, N. (2020). Stem-leaf segmentation and phenotypic trait extraction of maize shoots from three-dimensional point cloud. *ArXiv*. https://arxiv.org/abs/2009.03108

Zhu, X.-G., Long, S. P., & Ort, D. R. (2010). Improving Photosynthetic Efficiency for Greater Yield. *Annual Review of Plant Biology*, *61*(1), 235–261. https://doi.org/10.1146/annurev-arplant-042809-112206

Zou, H., Lu, H., Li, Y., Liu, L., & Cao, Z. (2020). Maize tassels detection: A benchmark of the state of the art. *Plant Methods*, *16*(1), 1–15. https://doi.org/10.1186/s13007-020-00651-z

Zuber, M. S. (1973). Registration of 20 Maize Parental Lines (Reg. Numbers PL 18 to 37). *Crop Science*, *13*(6), 779–780. https://doi.org/10.2135/cropsci1973.0011183x001300060085x

CHAPTER 5. GENERAL CONCLUSION

5.1. Conclusions

In this study, 3D machine vision and convolutional neural networks (CNNs) were utilized to advance the research of high-throughput robotic plant phenotyping both under controlled environments and in field conditions. Firstly, a non-destructive 3D scanning system and a graph-based 3D skeletonization algorithm were developed for indoor plant phenotyping of sorghum plants. Secondly, a customized stereo camera was developed and utilized to characterize organ-level phenotypic traits of sorghum and maize plants.

A low-cost machine vision system based on a commodity depth camera was developed to automated plant architectural traits characterization of sorghum plants grown under controlled environments. A ToF camera was mounted on a vertical track linear actuator to take sequential side-view images of sorghum plants at multiple developmental timepoints. The system allowed for cross-registration between the multi-level side-view point cloud by using the Iterative Closest Point (ICP) algorithm. To address the challenges caused by overlapped leaves and tillers, a 3D skeletonization algorithm was implemented by analyzing the 3D point cloud using a graph-based approach. The skeletonization algorithm was created by slicing the point cloud along the vertical direction, and then linking the connected Euclidean clusters between adjacent layers. Based on the structural clues of the sorghum plant, heuristic rules were implemented to separate overlapping tillers and individual leaves. Multiple parameters were obtained from the skeleton and the reconstructed point cloud, including plant height, stem diameter, leaf angle, and leaf surface area. The system-derived traits were highly correlated ($r = 0.81 \sim 0.94$) with the corresponding ground truth. Additionally, both the total leaf area and stem volume showed a great potential for wet and dry biomass prediction ($r = 0.82 \sim 0.98$). Consequently, depth imaging provides an efficient and economical solution for plant architecture characterization and phenotypic data extraction.

A customized stereo imaging module, namely PhenoStereo, was developed for field-based plant phenotyping. Integrated with a strobe lighting system, the PhenoStereo is capable of capturing high-quality and high-sharpness stereoscopic images under various environmental conditions (direct sunlight, backlighting conditions, shadows, wind conditions). Additionally, the camera module features onboard storage that performs image acquisition at a maximum rate of 14 frames per second (FPS). The PhenoStereo was deployed on a 4-wheel custom-built ground vehicle to collect side-view images of sorghum plants. Au automated image processing pipeline was developed to reconstruct the surface models and segment individual plant stems for quantifying stem diameter. A novel local feature descriptors *Cylindricity* was defined as a strict tensor to discriminate stem and non-stem points. The sorghum stems were modeled as cylindrical shapes by firstly find the cylindrical axis direction and then detect a circle in a 2D plane. The correlation coefficient (r) between the image-derived and ground truth measurements of stem diameter was 0.97 with an RMSE of 1.39 mm, surpassing the best values reported in the literature. The proposed method demonstrated that with proper customization stereo vision is feasible for 3D-based plant phenotyping under field conditions.

The PhenoStereo was also utilized to characterize leaf angle of maize plants in the field. Multiple tiers of PhenoStereo cameras were mounted on PhenoBot 3.0, a robotic vehicle designed to traverse between pairs of agronomically spaced rows of crops, to capture side-view images of maize plants in the field. A unique image processing pipeline named AngleNet was developed for leaf angle detection and characterization. The pipeline first utilized a CNN-based model to detect regions around leaf collars and three keypoints in each region in 2D images. The detection results in 2D images were back-projected to the 3D space to compute the actual leaf angle from reconstructed 3D models. The results illustrated that the trained CNN model could accurately detect regions of interest and keypoints for leaf angles with various leaf orientations and under different illumination and occlusion conditions. The mean and standard deviation of the angle errors in 2D images were 1.94° and 1.67°, respectively. The leaf angle estimates obtained using AngleNet were highly correlated ($r > 0.87$, RMSE $\leq 4°$) with in-field measurements. The AngleNet pipeline was used to explore the variations in leaf angle between two inbred maize lines: B73 and Mo17. The results of the distribution of leaf angles throughout the canopy were consistent with a previous report (Dzievit et al., 2019). Our study demonstrates the feasibility of using stereo vision to investigate the distribution of leaf angles in maize plant under field conditions.

5.2. Future Work

Plant phenotyping under controlled environments provides an optimal condition for image acquisition. Therefore, many studies have been conducted to analyze the morphological traits at the individual plant and organ scales. Under controlled environments, holographic images of plants can be obtained by multi-view imaging systems equipped with LiDAR (Wu et al., 2019), ToF cameras (McCormick et al., 2016), and high-resolution cameras (Sandhu et al., 2019; Wu et al., 2020). With one or more cameras rotating around the target plant, a holographic 3D surface model with higher accuracy and more details could be reconstructed. In addition to a high-fidelity surface model, accurate localization of objects in images is another essential aspect of high-performance phenotyping. The latter can be advanced by artificial intelligence (AI) using state-of-the-art CNNs. Hence, the future direction should focus on developing a multi-view 3D reconstruction system combined with AI for object detection and segmentation, where the challenges caused by leaf overlapping, occlusions, and homogenous texture could be alleviated.

Acquiring high-quality and high-sharpness images in the field is challenging due to variable lighting, wind conditions, plant occlusions, and complex background. Stereo vision combined with a strobe lighting system could potentially alleviate these challenges. Based on our experience, customized stereo cameras could produce more accurate point cloud under field conditions compared to other 3D sensors such as ToF cameras, LiDAR, and laser scanners. The performance of the PhenoStereo could be further improved by optimizing the light distribution pattern of the strobe lighting system. Currently, the strobe light is distributed in a circular pattern, which provides consistent lighting in the center of an image. The case is optimal for quantifying traits such as stem diameter and leaf angle from partial objects (e.g., partial stems, leaf collars). However, to measure traits such as leaf area and leaf length, the working area of the strobe lighting system should be larger, and the light pattern should be uniformly distributed in an image. This could be improved by modifying the geometrical distribution and tilting angle of the LEDs around the camera pair. Additionally, the field of view of the whole system could be further enlarged by increasing the number of cameras.

Generating high-fidelity 3D models is essential for accurately measuring plant architectural traits from 3D models. To achieve high-quality plant surface models, more advanced stereo matching algorithms could be used to improve the quality of disparity maps. For example, the performance of SGBM could be further improved by integrating edge-preserving techniques such as 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. Based on the standard benchmarking dataset, such as from the KITTI and the Middlebury, AI-based approaches usually have higher stereo matching accuracy than traditional methods (Hamid et al., 2020). For example, the MC-CNN-acrt network (Zbontar and LeCun, 2016), which uses an eight-layer CNN to produce matching costs for disparity map, performs better than any other published methods on three public datasets including KITTI 2012, KITTI 2015, and Middlebury dataset. There are some other end-to-end networks such as GC-Net (Kendall et al., 2017), PSMNet (Chang and Chen, 2018), iResNet (Liang et al., 2017), CRL (Pang et al., 2017) also achieved state-of-the-art results on public stereo benchmark dataset. CNN-based stereo matching has shown superior performance in accuracy and speed and the potential to overcome difficulties related to low texture areas, ill-posed regions, and near boundary regions.

The trend for solving computer vision problems in plant phenotyping using AI or machine learning becomes more apparent in recent years (Jiang & Li, 2020). The CNNs have demonstrated superior robustness in object detection and segmentation tasks. AI-based techniques can be applied on the side-view 2D images of maize and sorghum plants in the field to detect objects such as individual plants, leaves, stems, tassels, panicles, ears, and some other partial objects such as plant skeletons, leaf collars, branches, and internodes. Additionally, the CNNs also show a great potential for plant parts segmentation and architectural trait extraction in 3D space (Guo et al., 2021; Li et al., 2018; Qi et al., 2017). Therefore, it is worth exploiting CNN-based segmentation in both 2D images and 3D point cloud, which could potentially solve the challenges in plant phenotyping caused by diminished image texture, homogeneous plant surface, and dense canopy.

References

Chang, J. R., & Chen, Y. S. (2018). Pyramid stereo matching network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5410-5418). https://doi.org/10.1109/CVPR.2018.00567

Dzievit, M. J., Li, X., & Yu, J. (2019). *Dissection of Leaf Angle Variation in Maize through Genetic Mapping and Meta-Analysis*. 1–12. https://doi.org/10.3835/plantgenome2018.05.0024

Guo, M. H., Cai, J. X., Liu, Z. N., Mu, T. J., Martin, R. R., & Hu, S. M. (2021). Pct: Point cloud transformer. *Computational Visual Media*, *7*(2), 187-199. https://doi.org/10.1007/s41095-021-0229-5

Hamid, M. S., Abd Manap, N., Hamzah, R. A., & Kadmin, A. F. (2020). Stereo matching algorithm based on deep learning: A survey. *Journal of King Saud University-Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2020.08.011

Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A., & Bry, A. (2017). End-to-end learning of geometry and context for deep stereo regression. In *Proceedings of the IEEE international conference on computer vision* (pp. 66-75). https://doi.org/10.48550/arXiv.1703.04309

Jiang, Y., & Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: a review. *Plant Phenomics*, *2020*. https://doi.org/10.34133/2020/4152816

Li, Y., Bu, R., Sun, M., Wu, W., Di, X., & Chen, B. (2018). Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems, 31*.

Liang, Z., Feng, Y., Guo, Y., Liu, H., Qiao, L., Chen, W., ... & Zhang, J. (2017). Learning deep correspondence through prior and posterior feature constancy. *arXiv preprint arXiv:1712.01039*, *7*(8). https://doi.org/10.48550/arXiv.1712.01039

McCormick, R. F., Truong, S. K., & Mullet, J. E. (2016). 3D sorghum reconstructions from depth images identify QTL regulating shoot architecture. *Plant physiology*, *172*(2), 823-834. https://doi.org/10.1104/pp.16.00948

Pang, J., Sun, W., Ren, J. S., Yang, C., & Yan, Q. (2017). Cascade residual learning: A two-stage convolutional neural network for stereo matching. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 887-895). https://doi.org/10.1109/ICCVW.2017.108

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 652-660).

Sandhu, J., Zhu, F., Paul, P., Gao, T., Dhatt, B. K., Ge, Y., ... & Walia, H. (2019). PI-Plat: a high-resolution image-based 3D reconstruction method to estimate growth dynamics of rice inflorescence traits. *Plant Methods*, *15*(1), 1-13. https://doi.org/10.1186/s13007-019-0545-2

Wu, S., Wen, W., Xiao, B., Guo, X., Du, J., Wang, C., & Wang, Y. (2019). An accurate skeleton extraction approach from 3D point clouds of maize plants. *Frontiers in plant science*, *10*, 248. https://doi.org/10.3389/fpls.2019.00248

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*. https://doi.org/10.34133/2020/1848437

Zbontar, J., & LeCun, Y. (2016). Stereo matching by training a convolutional neural network to compare image patches. *J. Mach. Learn. Res.*, *17*(1), 2287-2318.

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). https://doi.org/10.1109/ICCV.2015.49