Crop Stem Width Estimation in Highly Cluttered Field Environment

Anwesa Choudhuri
anwesac2@illinois.edu
Girish Chowdhary
girishc@illinois.edu

University of Illinois
Urbana-Champaign
Illinois, USA

Abstract

We present two methods for estimation of crop stem width on small mobile robots. Stem width is an important phenotype needed by breeders and plant-biologists to measure plant growth, however, its manual measurement is cumbersome, in-accurate, and inefficient. The presented methods use a common image processing core that is designed to extract the foreground in the presence of significant leaf and stem clutter, view of other rows, and varying lighting, from a side-facing USB camera mounted on a small mobile robot. Using the extracted foreground, one approach uses estimates of robot velocity from wheel encoders and dense optical flow to estimate depth, while the other employs filtering of the LIDAR 2-D point cloud to estimate the depth. Both methods have been exhaustively validated against available hand-measurements on biomass sorghum (Sorghum bicolor) in real experimental fields. Experiments indicate that both methods are also applicable to other crops with cylindrical stems without significant modifications. The width estimation match on sorghum is 92.5% (using vision) and 98.2% (using vision and LIDAR) when compared against manual measurements by trained agronomists. Thus, our results clearly establish the feasibility of using small robots for stem-width estimation in realistic field settings. Furthermore, the techniques presented here can be utilized for automating other phenotypic measurements.

1 Introduction

Plant phenotyping is the quantification of the effects of genotype differences (i.e., differences in the genetic makeup) and the environment on the exhibited phenotype (i.e., the plant appearance and behavior) [27]. According to the Food and Agriculture Organization of the United Nations, large-scale experiments in plant phenotyping are a key factor in breeding better crops that are needed for feeding a growing population and providing biomass for energy, while using less water, land, and fertilizer. The need for large scale, more comprehensive, and efficient phenotyping has become ever more pressing recently due to a constantly evolving climate [20] and changing demographic in rural areas. However, current phenotyping methods are mostly limited to manual measurements out in the field, which is labor intensive, time consuming, and lacks sufficiency and accuracy. This has created a so-called phenotyping bottleneck in agricultural productivity increase.

1.1 Background on automated plant phenotyping

Over the past few years, several attempts have been made to automate the process of plant phenotyping using a wide range of sensors involving multispectral and hyperspectral remote sensing, thermal infrared imaging, fluorescence imaging, 3D imaging, tomographic imaging and imaging in visible light. Li et al. [17] provide an extensive survey of the aforementioned methods, their applications, advantages and limitations.
Visible imaging is a practical, energy efficient and cost effective way to measure several plant phenotypes. Many recent approaches ([9], [15], [21], [13], [26], [5], [30], [7]) try to model plants using imaging techniques and 3D reconstruction. However, these approaches have been tested under simulated environment or under extensively

(a) (b) (c)

Figure 1: (a) Image of rosette plant from the dataset given by Minervini et al. [22]. (b) Images acquired from real robot in actual field condition with high level of clutter, varying sunlight, motion blur, and occlusion. Images obtained by robots are very different from the datasets available. (c) Width measurement using vernier calipers is cumbersome and inaccurate

monitored environments, such as a green house. The aforementioned methods and algorithms have not been implemented in actual agricultural fields where the level of uncertainty is very high due to changes in lighting conditions during different times of the day and during different seasons, variation of plant color and size across different growth stages, background clutter, and numerous other factors. Hoyos-Villegas et al. [14] and Chen et al. [6] try to do experiments in field conditions and use digital imaging to assess soyabean and sorghum respectively: Hoyos-Villegas et al. [14] develop a hand-held digital imaging tool to assess soybean yield, while Chen et al. [6] use UAS based RGB imagery to locate plant centers of sorghum in a field but achieves an accuracy of only 64% to 66%. In contrast, our tests are performed in harsh and variable field conditions, yet the accuracy level is over 90%. Furthermore, our results are designed and evaluated on ground robots, which need to deal with harsher conditions, but are more desirable for high throughput phenotyping as opposed to UAS, since they have a far closer and more detailed under-canopy view of plants [23].
Although there is a huge potential of deep learning and computer vision in plant phenotyping, it has become clear that the challenges arising in plant phenotyping differ significantly from the usual tasks addressed by the computer vision community [27]. In context of robotic phenotyping, one huge challenge is the lack of available labeled and curated datasets for training deep networks in realistic field conditions. Minervini et al. [22] provide a dataset of potted rosette plants over multiple growth stages where each leaf segment of the plants are labeled with a different color [19]. Many recent works have used this dataset to achieve high accuracy in leaf counting and segmentation tasks ([1],[31],[28],[8]). Giuffrida et al. [12] augment this dataset by generating realistic images of rosette plants using generative adversarial networks. Pound et al. [25] provide a new dataset for wheat spikes, analyze the spikes and count them. However, the conditions in this dataset differ significantly from those in the field (e.g. Figure 1.1(a),(b)), with field obtained data from a moving robot having a high level of clutter, varying sunlight, motion blur, and occlusions. The significant contribution of our work is in creation of pipelines that can work directly with robot obtained data in field conditions. Our elaborate algorithms for filtering useful content from noisy robot-obtained field images will lead to datasets that could feed future machine learning pipelines.

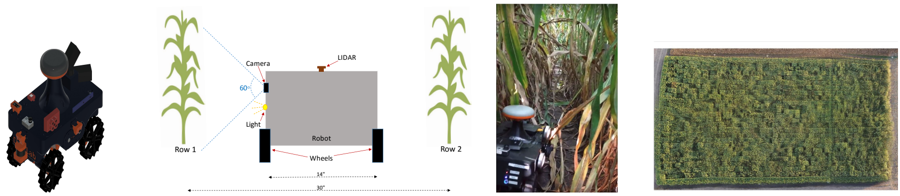
1.2 State of the Art in Stem Width Estimation

Stem width of fuel plants is an important phenotype that determines the plant's biomass content. In spite of its importance, there are currently no efficient in-field practices for stem width measurement. The current practice involves trained agronomists manually going out into fields and measuring the stems using vernier calipers (see figure 1.1(c)). This technique is slow, inaccurate, hazard-prone, and highly labor intensive. Jin and Zakhor [16] propose an algorithm to estimate stem width from 3D point cloud data collected by a robot equipped with a depth sensor (ToF sensor). Baharav et al. [2] use 2 infrared cameras mounted on a robot and applies image processing techniques to estimate height and width of plants. However, none of the existing width estimation algorithms have yet been validated in field settings for accuracy and validity under high clutter and changing field conditions.

1.3 Contribution

Our main contribution is two new pipelines for crop stem width estimation under high clutter in an agricultural field using low cost sensors mounted on a small mobile robot called TerraSentia, that can traverse autonomously through crop rows. Details of the robot and sensor are provided in section 2. One of our contribution is an image processing pipeline to extract the foreground in the presence of significant leaf and stem clutter, view of other rows, and varying lighting. Using this, one pipeline uses the ratio between estimated robot velocity from wheel encoders [1] and pixel velocity from dense optical flow to estimate depth. Our contribution here is in adapting an optical flow based approach (3.2.1) for phenotyping in cluttered and unstructured field conditions. Structure from motion can

[1]Note that GPS does not work well under crop canopy due to multi-path and attenuation errors, hence, encoder velocity is an acceptable estimate of robot speed, especially at slow speeds when the wheels do not slip excessively.



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

Figure 2: (a) CAD drawing of the phenotyping robot TerraSentia that was used for data acquisition; (b) Front View of the Robot in a 76 cm row, showing placement of the camera, light and LIDAR. Field of view of the camera is 60°. (C) TerraSentia moving autonomously through a 30 inch wide, heavily cluttered sorghum row. Autonomous navigation is performed by path estimation from point cloud data obtained from the 2D LIDAR (using Algorithm 3 provided in the supplementary material). (d) Aerial view of 80 acres (0.32 sq. km) sorghum field (Maxwell Field, Savoy, IL, USA), consisting of 960 sorghum plots of different genotypes of Sorghum.

be used for 3D reconstruction of plants in cluttered field environment as an extension to this work. Our second pipeline employs filtering of the LIDAR point cloud to estimate the robot-to-row depth from highly noisy LIDAR point-cloud (3.2.2). This is another contribution of our work, and our results demonstrate that this filtering method can also be used for robot navigation.

A significant contribution of our work is in the exhaustive validation in biomass sorghum fields (elaborated in 4). We argue that our measurements could potentially be more accurate than those of the human agronomist, especially given that the agronomist can only measure limited plants and only in a few places, whereas our algorithm can be more exhaustive.

The algorithms presented here are quite general in nature due to their reliance on fundamental principles of machine vision. As such, they can be utilized with little modification on other plants, as demonstrated by our experiments in corn (Zea mays) and hemp (Cannabis) fields (the data will be released upon acceptance). Therefore, our results establish the feasibility of using small autonomous robots for phenotyping in realistic field environments.

The outline of the paper is as follows. Descriptions of the robot, sensors and some information about the sorghum plots are provided in section 2. Section 3 provides detailed explanations of the algorithms. The test results are summarized in section 4, followed by challenges and limitations of the work in section 5 and conclusion in 6.

2 Experimental Set-Up

2.1 Robot Description

A lightweight(<7 kg), ultra-compact, 3D printed, autonomous field phenotyping robot (called TerraSentia) is used for data acquisition (Figure 2(a),(b)). (Details of the robot are provided in the supplementary file). The lightness and thoughtful construction of the wheels prevent permanent damage to any plant parts, even if the robot runs over them accidentally. The compactness allows the robot to easily traverse between narrow crop rows, especially in corn, hemp, and sorghum, where 76cm or greater row spacing is common. Each wheel of the robot is powered by a separate motor with encoders and averaged encoder values provide a reasonable estimate of the robot speed. We utilized the robot at a speed of about 0.4 m/s. At this speed, the robot covers a row of a single 3m by 3m plot of a crop variety in under 10 seconds, and can cover several plots in a reasonable amount of time. Furthermore, at this speed, motion blur was not found to be significant on the camera (ELPUSBFHD01M, USA) used on the robot. Higher frame rate cameras could enable increasing speed of the robot.

2.2 Robot Navigation

The robot system has been designed for reliable autonomous navigation under cluttered canopies using only a 2-D Hokuyo UTM-30LX LIDAR sensor as the primary sensor for navigation. The primary reliance on LIDAR comes from the fact that GPS is highly unreliable under dense crop canopy due to multipath errors and signal attenuation. Figure 2 (c) shows the robot traversing autonomously with the help of path estimation by the LIDAR.

2.3 Camera and Light Source

As the robot traverses through crop rows, video data is acquired at 90 frames per second with a frame resolution of 640×480 pixels by the robot's camera. The acquisition is performed with a low-cost digital monocular RGB camera (ELPUSBFHD01M, USA) mounted on the side of the robot chassis. The camera has a field view of 60°.

A common inexpensive LED light source, having a 3000K color temperature and providing 60 lumens of light, is attached on the robot chassis near the camera position to ensure ample brightness under the dark sorghum canopy. This is done to prevent the camera firmware from increasing the exposure time which could cause excessive motion blur. We had no control over the camera firmware. The placements of the camera and light are shown in Figure 2(b).

2.4 Data Acquisition

Video data, encoder readings (for instantaneous robot velocity estimation) and LIDAR point cloud data (for lateral distance estimation) are acquired at 90 fps, 5 fps and 20 fps respectively. GPS data is also logged, however, GPS accuracy varies widely under canopy, hence GPS data is not used. There was no inertial measurement unit on the robot we utilized. All our results are in an offline setting with data retrieved using wi-fi, however, high speed and low computational requirements of the presented algorithms indicates that they could be utilized onboard the robot in the future. In either case, there is little value lost by doing off-board estimation, since the data has to be retrieved for other purposes any way. Python 2.7.12 with OpenCV 2.4.9.1 have been used for development of all codes. We plan to release code upon acceptance.

2.5 Description of the Sorghum Fields

All experiments on sorghum have been performed on the 80 Acre (0.32 sq. km) Maxwells Field, Savoy, Illinois, during August to November, 2017. The field consists of 960 Sorghum plots of size 3m$\times$3m and 76 cm row spacing. Figure 2(d) shows an aerial image depicting the marked height difference due to the different genotypes in each plot.

3 Algorithmic Framework

We design two algorithmic approaches for robust estimation of crop stem width under highly uncertain field conditions. The algorithms have been divided into two phases. Phase 1 is the same for both algorithms, and involves a common image processing algorithm for foreground extraction (algorithm 1). Phase 2 consists of depth estimation using structure from motion and LIDAR point clouds for both approaches respectively. The framework has been summarized in figure 3 and described in details in the following sub-sections.

3.1 Phase 1: Foreground Extraction

A fixed sized window is defined at the left side of each video frame and only that region is processed to extract a stem boundary if present. The choice of this size is based on two assumptions valid for most crop fields: (1) The stem width does not exceed 4 inches, (2) Stems never come closer than 3 inches to the camera lens without blur. This window size has worked for all our validation tests, and hence we avoid using a variable-sized window to avoid unnecessary and redundant calculations. Placement of the window is on a side instead of the center because video frames sometimes blur towards the center due to small row spacing, in spite of the high frame rate (90 fps) and additional lights (section 2). We have placed the window at the left side, but choosing the right side is also acceptable. The placement of the window is shown in (fig. 4).

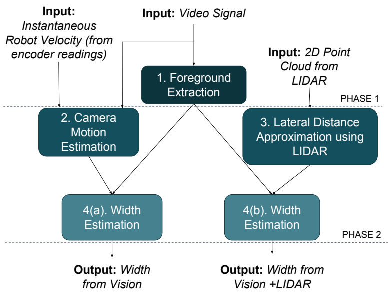


Figure 3: Algorithmic Framework, (Refer to supplementary material for pseudo-code: (1) Foreground Extraction (algorithm 1); (2) Camera Motion Estimation from SFM (algorithm 2); (3) Lateral Distance Estimation Using LIDAR (algorithm 3); (4) Width Estimation using LIDAR and SFM (algorithm 4))

The window, placed in each frame of the raw video, is processed to find out if an un-occluded part of the stem is visible. The rest of the video frame does not need to be utilized, and this is done to reduce computational overhead. Discarding the rest of the frame does not cause any valuable information loss since the robot traverses through the entire row visiting each plant one by one, so all parts of the frame pass through the chosen window at some point in time. Approach of multiple windows is avoided to reduce the chance of double counting of the same plant multiple times, which distorts the estimated width distribution across the plot.
Figure 5 shows the sequential processing of the window obtained for foreground extraction. Algorithm 1 provides an outline of the same, which are described in the following subsections. The pseudo-codes have been provided in algorithm 1 of the supplementary material.

Figure 4: Window (marker with cyan) placed at the left corner of the video frame. Center of frame can often be blurred.

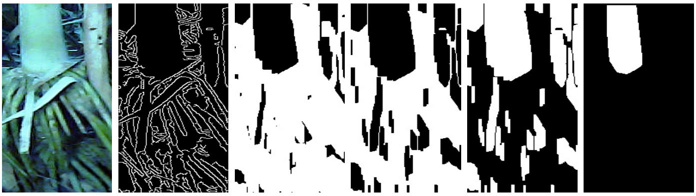


Figure 5: Sequential steps for foreground extraction from cropped window of current video frame: (a) Unprocessed window f_n with high clutter; (b) Edges of (a) after Canny edge detection; (c) Dilation of (b); (d) Erosion of (c); (e) Inversion of (d); (f) Foreground extraction and smoothing by removing unwanted components (e) after CCL and convex hull approximation respectively. (Refer to algorithm 1 of the supplementary material for pseudo-code)

Algorithm 1 Foreground Extraction of cropped video frame from cluttered background. (Refer to supplementary material (algorithm 1) for pseudo-codes.

1: **procedure** EDGES(f_n) ▷ Function to return edges (f_e) from cropped video frame f_n

2: **procedure** MORPH(f_e) ▷ f_{e2} : Function to perform dilation, erosion, inversion on the edges(f_e) and return cluttered foreground (f_{e2})

3: **procedure** FOREGROUND(f_{e2}) ▷ Function to perform connected component labeling on f_{e2} to remove clutter, and to smoothen foreground mask. Returns clean and smooth foreground $fullMask$

3.1.1 Canny Edge Detection

Edge detection by Canny's technique [4] is adopted and illustrated in figure 5(b). We have adopted this technique because of it's easy availability, and robustness over other available edge detection techniques (like Sobel, Laplacian [3]).

3.1.2 Morphological Operations

Owing to noise and variable lighting conditions, the edges obtained in the previous step are often broken. We close the edges by using morphological dilation followed by erosion [3] and inverting the image (Figure 5(c),(d),(e)). A rectangular kernel of size 25×3 is used for dilation followed by a kernel size of 15×15 for erosion. Using a long rectangular kernel helps restore the horizontal width information; the vertical gradients become distorted, which are not of significance in this application. The kernel sizes have been selected empirically after trial on numerous plots of corn, sorghum and hemp, in different fields and different times of the day.

3.1.3 Components Removal

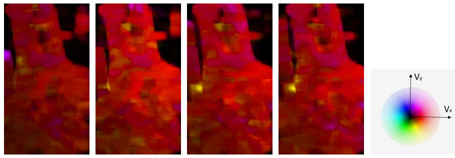
Connected component labeling (CCL) is performed for the extraction of stem from a cluttered image like (Figure 5(e)), performed using the skimage library in python ([11] and [32]). A labeled image is obtained in which each white component of Figure.5(e) is indexed differently. Using skimage we measure essential characteristics of each labeled component and remove the objects that do not have the desired characteristics of a stem. For example, a stem must have larger size than background clutter, higher eccentricity than undesired leaf and more erect orientation than the diverging branches. We require only the stem components which are cylindrical, and are projected as long rectangles on the video frames. If we approximate ellipse around them, the eccentricity of such ellipses are high(>0.8). We also provide the constraint of orientation of $40°$ on the left and right. These three constraints (size, eccentricity and orientation) remove most of the background clutter and non-stem components like leaves or other objects like sign boards or shoes of a person walking. Figure5(f) shows the cleaned up image after this step.

3.1.4 Polygon Approximation

A rough mask of the plant stem is obtained after the above process that is clean and devoid of unnecessary components. To smoothen the mask, we find a convex hull of these 2D point sets using Sklansky's algorithm [29]. This is shown in Figure 6.

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

Figure 6: Foreground smoothing: (a)&(c) Rough foreground after CCL, (b)&(d) Smooth foreground after approximating convex hull around the object of interest



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

Figure 7: Color-coded visualization of dense optical flow using Farnebäck [10] algorithm: (a) to (d) Optical Flow for 5 consecutive windows. Robot moves in the positive x direction. Brighter means higher velocity, (e) Color Code for directions. (Refer to algorithm 2 of the supplementary material for pseudo-code)

3.2 Phase 2

3.2.1 Camera motion estimation

Structure from Motion involves the three main steps of (1) extraction of features in images and matching these features between images, (2) camera motion estimation (e.g., using relative pairwise camera positions estimated from the extracted features), and (3) recovery of the 3D structure using the estimated motion and features (e.g., by minimizing the so-called re-projection error) [24].
We require steps (1) and (2) in our application to determine average pixel velocity of the clean foreground in each frame. Step (1) is performed using Gunnar Farneback's algorithm [10]. This algorithm computes the dense optical flow for all points in the window and gives a 2 channel array with optical flow vectors (V_x, V_y). Figure 7 shows the color-coded visualization of the calculated optical flow. Brighter objects have a higher velocity component and hence are closer to camera. The pseudo-code has been provided in algorithm 2 of the supplementary material.
Approach of dense optical is adopted instead of sparse optical flow, that involve tracking of feature points across video frames (like Lucas Kanade algorithm [18]), because of the lack of distinctively different points on the stems and leaves in a sorghum field.

3.2.2 Lateral Distance Estimation Using LIDAR

Figure 8(a) shows the point cloud (in black circles) provided by the 2D LIDAR as output. Although sensor range is up to 30m, meaningful readings do not surpass a few meters, because row spacing is below 1m. Frequent occlusion by leaves and weeds interfere with measurements from ahead of the lane, hence, the points too close to the robot are discarded as clutter by leaves. The remaining points are shown by red crosses in 8(a) and (b)). Appropriate lines are fitted through the chosen points by a weighted mean of the point distances to the line of traversal of the robot. This is shown in 8(c). The lateral LIDAR to plant-row distance is calculated in the desired side, where the camera is pointed, and lateral distance from the camera to the LIDAR is subtracted which gives us the desired D (camera to row distance) useful for width estimation. The steps for estimating D are summarized in algorithm algorithm 2 of the supplementary material.

3.3 Width Estimation from SFM and LIDAR data

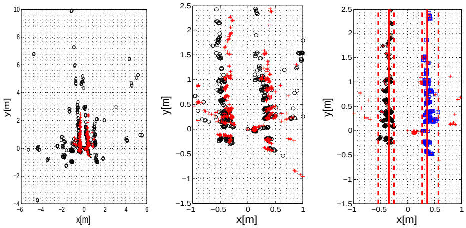
A clean white foreground mask on black background, which has the same length and breadth as the fixed window, is obtained after section 3.1, which is the mask for the desired stem boundary. N imaginary horizontal lines are drawn on the mask image and for each line, the number of white pixels are recorded. This gives us the width in pixels (W_{P_i}s) at N different locations of the stem (as shown in Figure 9). We experimentally chose the value of N to be 8 here. The actual width estimation process from pixel width is described in the following sub-sections. (Refer to algorithm 4 of the supplementary material for pseudo-code)

3.3.1 From SfM

To obtain the actual stem width W_{S_i}, equations 1, 2 and 3 are followed. V_X is calculated from section 3.2.1, V_R is the instantaneous velocity of the robot obtained from encoder readings. The width (W_S) obtained for a particular window is the average of all the W_{S_i}.

$$R = \frac{V_R}{V_X} \quad (1) \qquad W_{S_i} = W_P \times R \quad (2) \qquad W_S = \frac{1}{N}\sum_{i=1}^{N} W_{S_i} \quad (3)$$

This approach of using SfM obviates the need to use complex sensors like the realsense or stereo camera to estimate the depth information.



(a) (b) (c)

Figure 8: Lateral Distance Approximation. (a) LIDAR point cloud shown by black circles, note the significant number of outliers. Red crosses indicate the chosen points after rejection of outliers, (b) Zoomed view of (a), (c) Lane Approximation as shown by fitted red lines (Refer to algorithm 3 of the supplementary material for pseudo-code)

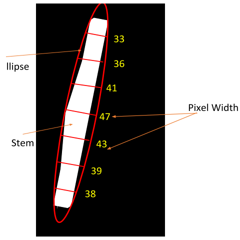


Figure 9: Pixel Width (W_{P_i}) at N different places (N=8) of a sorghum stem

3.3.2 From LIDAR

The instantaneous distance from the camera to the crop row under consideration (D) is obtained according to section 3.2.2 The width from LIDAR W_L is calculated according to eqns 4 and 5. ‘F’ refers to the focal length of the camera.

$$W_{L_i} = W_P \times D \div F \tag{4}$$

$$W_L = \frac{1}{N} \sum_{i=1}^{N} W_{L_i} \tag{5}$$

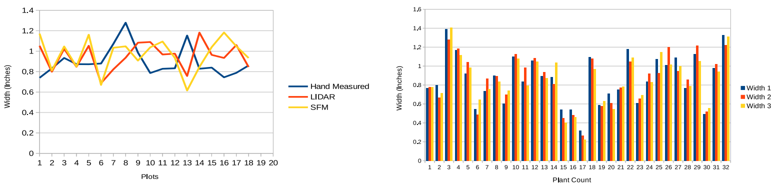
W_S and W_L are the outputs from two algorithmic approaches that we adopted. The results obtained after exhaustively validating these estimations are expressed in the next section.

4 Results

A significant contribution of our work is in the exhaustive validation of the presented algorithms in biomass sorghum (Sorghum bicolor (L.) Moench) in real fields. All experiments have been performed near the last growth stage of sorghum, when the clutter and leaf occlusion is the highest among all growth stages.

4.1 Expt 1: Comparison with agronomists

The algorithms have been compared against all available manually measured plots in the 80 Acre (0.32 sq. km) sorghum experimental field (Maxwell fields at Savoy, Illinois). Each 3 $\times$ 3 meter plot consists of roughly 50 plants. A trained agronomist (independent from the authors) used industry-standard practices to measure the average stem-width of 20 plots dispersed across the field. The agronomist chose in each plot 3 representative plants, and made one manual measurement with vernier calipers from each of these 3 plants. This approach was designed to enable the agronomist to accumulate a reasonable amount of data from large fields within cost and time constraints. The industry practice is to use the average of these 3 readings to represent the average stem width for that plot. On the other hand, our robot traversed through those 18 plots and attempted to measure stem width at multiple locations on every plant in the plot. The comparison is shown in Figure 10(a). The percentage match of our algorithms with the hand measured width by the agronomists, considering all plots, is 78% using LIDAR and 76% from SFM. This



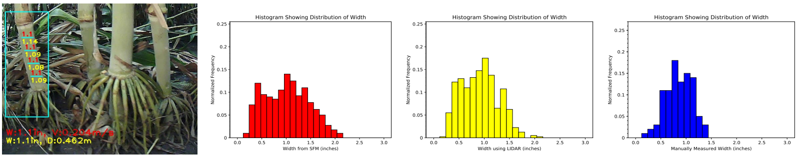
(a) (b)

Figure 10: (a) Distribution of Manual Measurements, LIDAR Width and SFM Width across 18 Plots of sorghum in Maxwell’s field. (b) Variation of manual measurement per plant in plot 17MW0159. 3 measurements have been taken per plant for all plants in the plot. Standard Deviation: 0.256 inches

disagreement however is not surprising, since the sparse manual measurements of the agronomist are not reflective of the ground truth, nor do they represent the true nature of the width distribution of plants. Manual measurements are limited by cost and time considerations. They do not take into account the fact that the width of the plants varies along its length significantly, so a single measurement does not reflect the true plant width. Furthermore, the cross section of the stem is elliptical, not circular, therefore, the placement of vernier calipers affects the measurement. Figure 10(b) depicts the amount of width variation in a single plot (see Section 4.2). This lack of accuracy and rigor due to high cost of trained manual labor is precisely the reason the industry is looking towards robotic high-throughput phenotyping. Therefore, to evaluate our algorithm against a true representation of the stem width, we performed Experiment 2.

4.2 Expt 2: Comparison with extensive hand measurements

To address the limited agronomist obtained measurements, extensive manual measurements were performed on a representative plot (17MW0159) in the sorghum field. 3 measurements were taken from different lengths of each plant from one row of the plot which consisted of 32 sorghum plants. Figure 10(b) show the variation in the 32×3 measurements.



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

Figure 11: Width estimation (in inches) for sorghum: (a) A typical video frame for plot 17MW0159 with window marked in cyan. W_{S_i} (red) and W_{L_i} (yellow) are noted inside window. W_S (red) and W_L (yellow) are placed at the bottom left corner. V: Instantaneous robot velocity, D: Robot to plant distance measured by LIDAR); (b) Distribution of Width from SfM Across all frames in the video; (c) Distribution of width from LIDAR across all frames in the video; (d) Distribution of hand measured width for all plants in the plot

The standard deviation of such a distribution is 0.256 inches, clearly showing that averaging only 3 measurements per plot (as in Section 4.1), which is the current practice, is not very accurate. Even these measurements however, do not represent the "ground truth", however, they are the best available set of data against which we could compare. Table 4.2(b) shows that the robot based algorithms match with the exhaustive hand measurements at 91% when structure from motion is used, and 98% when vision and LIDAR are used, which is under the 8% tolerance set by a federal oversight agency. Figure 11 (a) shows a processed video frame for plot 17MW0159 as the robot traversed through the plot. The processing window after cropping the video frame is marked with cyan. The red and yellow colors indicate width using structure from motion and LIDAR data respectively. The values inside the window are the instantaneous width values (Red: W_{S_i} and Yellow: W_{L_i}) corresponding to different 'i's as described in section 3.3, whereas the values at the bottom left corner are the average values for each method (Red: W_S and Yellow: W_L). All values are in inches. Figure 11 (b) shows the distribution of width for presented algorithms and hand measured ground truth. Table 4.2(a) shows mean and variance of such a measurement.

	LIDAR	SFM	Manual
Mean (")	0.84528	0.93360	0.86014
Variance (")	0.13247	0.20301	0.06623

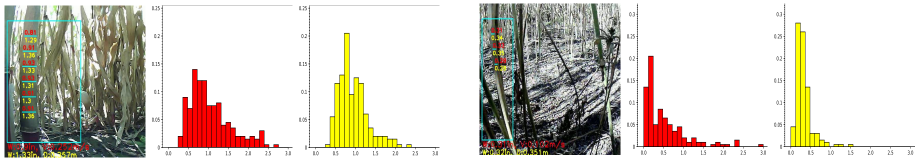
(a)

	Manual	LIDAR	SFM
Width (")	0.86014	0.84528	0.93360
% Match		98.27	91.4587

(b)

Table 1: (a) Comparison of Mean and Variances of Width Estimation from Presented Algorithms and Hand Measured Values for plot 17MW0159. (b) Comparison of Manual Measurement with results obtained from the presented algorithms for plot 17MW0159

The results obtained after comparing the width estimation results with the average measurements from plot 17MW0159 are tabulated in Table 4.2(b). The robot speed estimates from the averaged encoder readings cause the lower accuracy from the optical flow approach as compared to the LIDAR approach, but the error is within acceptable range. Hence, we refrain from using costlier high precision encoders. The processing time is 0.04 secs per video frame.



(a) (b) (c)
Width for Corn

(d) (e) (f)
Width for Hemp

Figure 12: Width (inches) for corn ((a),(b),(c)) and hemp ((d),(e),(f)): (a),(d) Video frame with window marked in cyan. W_{S_i} (red) and W_{L_i} (yellow) are noted inside window. W_S (red) and W_L (yellow) are placed at the bottom left corner. V: Instantaneous robot velocity, D: Robot to plant distance measured by LIDAR). (b),(e) Distribution of Width from SfM across all video frames. x-axis: Width in inches, y-axis: Normalized frequency. (c),(f) Distribution of width from LIDAR across all frames in the video. x-axis: Width in inches, y-axis: Normalized Frequency

4.3 Demonstration of generalization to other crops

The presented algorithms have been extensively validated in sorghum across several plots consisting of different varieties of sorghum, at different times over a period of days and in different weather conditions (under bright sun and cloud cover). The results in all cases of Sorghum remain consistent. To further demonstrate robustness and generality, the algorithms have also been applied to other crops with cylindrical stems: corn and hemp, with no algorithmic modifications and a few changes in the parameters for edge detection. The results on corn and hemp data obtained at fields in parts of Illinois and Colorado are shown in Figure 12. The width estimates are within the expected ranges, however, we were not able to rigorously establish the ground truth on these crops due to lack of manual resources. Regardless, the results demonstrate that our algorithms, based on principles of machine vision, generalizes well across different crops with little or no change to the parameters.

5 Challenges and Limitations

In this section we discuss a few challenging situations that require future work. Figure 13 (a) shows a situation where even manual classification of the image for stem or leaf is difficult. The picture shows a leaf, having color, size, eccentricity, orientation and shape just like the stem. Hence the algorithm falsely detects this as a stem. Figure 13 (b) shows a stem almost completely occluded by a leaf in front, there's no way to detect the stem in this case with the sensors we have used. Figure 13 (c) shows bright sunlight entering through dense sorghum canopy, causing stems and leaves to be only partially illuminated. In this case, only the partial contour is taken into account, leading to faulty width estimation.

(a) (b) (c)

Figure 13: Hard problems, detection can fail under these situations:(a) A leaf looking like a stem (similar size, shape, eccentricity, orientation); (b) Stem occluded by leaf (c) Strong uneven sunlight

Training a machine learning algorithm could be one way to deal with such situations, but that could requires thousands of labeled frames of videos taken under extreme field conditions. Another drawback of using deep learning is that the algorithms becomes crop specific, thereby losing generality, unlike our approach. Regardless, within reasonable weather and crop cover, the results from our robot based method should yield far more accurate and rich data than manual measurements alone.

6 Conclusion

We have presented a general and computationally light algorithm for autonomous width estimation of crops with cylindrical stems under highly uncertain field conditions and exhaustively validated the results with rigorous ground truth measurements for sorghum. This is a significant step forward in field robotics for phenotyping applications. Our next step is to use these algorithmic approaches to develop a dataset specifically for field conditions, with masked labels for stems and leaves. This would enable the use of machine learning to initiate future work on important phenotype estimation tasks like leaf area, leaf angle,leaf count and further improved stem width.

References

[1] Shubhra Aich and Ian Stavness. Leaf counting with deep convolutional and deconvolutional networks. *CoRR*, abs/1708.07570, 2017. URL http://arxiv.org/abs/1708.07570.

[2] Tavor Baharav, Mohini Bariya, and Avideh Zakhor. In situ height and width estimation of sorghum plants from 2.5d infrared images.

[3] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer Vision in C++ with the OpenCV Library*. O'Reilly Media, Inc., 2nd edition, 2013. ISBN 1449314651, 9781449314651.

[4] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, Nov 1986. ISSN 0162-8828. doi: 10.1109/TPAMI.1986.4767851.

[5] Ayan Chaudhury, Christopher Ward, Ali Talasaz, Alexander G. Ivanov, Mark Brophy, Bernard Grodzinski, Norman P. A. Huner, Rajni V. Patel, and John L. Barron. Machine vision system for 3d plant phenotyping. *CoRR*, abs/1705.00540, 2017. URL http://arxiv.org/abs/1705.00540.

[6] Yuhao Chen, Javier Ribera, Christopher Boomsma, and Edward Delp. Locating crop plant centers from uav-based rgb imagery. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[7] Sruti Das Choudhury, Saptarsi Goswami, Srinidhi Bashyam, A. Samal, and Tala N. Awada. Automated stem angle determination for temporal plant phenotyping analysis.

[8] Andrei Dobrescu, Mario Valerio Giuffrida, and Sotirios A. Tsaftaris. Leveraging multiple datasets for deep leaf counting. *CoRR*, abs/1709.01472, 2017. URL http://arxiv.org/abs/1709.01472.

[9] Lingfeng Duan, Wanneng Yang, Chenglong Huang, and Qian Liu. A novel machine-vision-based facility for the automatic evaluation of yield-related traits in rice. *Plant Methods*, 7(1):44, Dec 2011. ISSN 1746-4811. doi: 10.1186/1746-4811-7-44. URL https://doi.org/10.1186/1746-4811-7-44.

[10] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th Scandinavian Conference on Image Analysis*, SCIA'03, pages 363–370, Berlin, Heidelberg, 2003. Springer-Verlag. ISBN 3-540-40601-8. URL http://dl.acm.org/citation.cfm?id=1763974.1764031.

[11] Christophe Fiorio and Jens Gustedt. Two linear time union-find strategies for image processing. *Theoretical Computer Science*, 154(2):165 – 181, 1996. ISSN 0304-3975. doi: https://doi.org/10.1016/0304-3975(94)00262-2. URL http://www.sciencedirect.com/science/article/pii/0304397594002622.

[12] Mario Valerio Giuffrida, Hanno Scharr, and Sotirios A. Tsaftaris. ARIGAN: synthetic arabidopsis plants using generative adversarial network. *CoRR*, abs/1709.00938, 2017. URL http://arxiv.org/abs/1709.00938.

[13] Mahmood R. Golzarian, Ross A. Frick, Karthika Rajendran, Bettina Berger, Stuart Roy, Mark Tester, and Desmond S. Lun. Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant Methods*, 7(1):2, Feb 2011. ISSN 1746-4811. doi: 10.1186/1746-4811-7-2. URL https://doi.org/10.1186/1746-4811-7-2.

[14] V. Hoyos-Villegas, J.H. Houx, S.K. Singh, and F.B. Fritschi. Ground-based digital imaging as a tool to assess soybean growth and yield. *Crop Science*, 54:1756–1768, 2014. doi: 10.2135/cropsci2013.08.0540.

[15] Mayuko Ikeda, Yoshitsugu Hirose, Tomonori Takashi, Yosuke Shibata, Takuya Yamamura, Toshiro Komura, Kazuyuki Doi, Motoyuki Ashikari, Makoto Matsuoka, and Hidemi Kitano. Analysis of rice panicle traits and detection of qtls using an image analyzing method. *Breeding Science*, 60(1):55–64, 2010. doi: 10.1270/jsbbs.60.55.

[16] Jihui Jin and Avideh Zakhor. Point cloud based approach to stem width extraction of sorghum. 2017.

[17] Lei Li, Qin Zhang, and Danfeng Huang. A review of imaging techniques for plant phenotyping. *Sensors*, 14(11):20078–20111, 2014. ISSN 1424-8220. doi: 10.3390/s141120078. URL http://www.mdpi.com/1424-8220/14/11/20078.

[18] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81, pages 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc. URL http://dl.acm.org/citation.cfm?id=1623264.1623280.

[19] M. Minervini, A. Fischbach, H. Scharr, and S.A. Tsaftaris. Plant phenotyping datasets, 2015. URL http://www.plant-phenotyping.org/datasets.

[20] M. Minervini, H. Scharr, and S. A. Tsaftaris. Image analysis: The new bottleneck in plant phenotyping [applications corner]. *IEEE Signal Processing Magazine*, 32(4):126–131, July 2015. ISSN 1053-5888. doi: 10.1109/MSP.2015.2405111.

[21] Massimo Minervini, Mohammed Abdelsamea, and Sotirios A. Tsaftaris. Image-based plant phenotyping with incremental learning and active contours. 23:35–48, 09 2014.

[22] Massimo Minervini, Andreas Fischbach, Hanno Scharr, and Sotirios A. Tsaftaris. Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognition Letters*, 81:80 – 89, 2016. ISSN 0167-8655. doi: https://doi.org/10.1016/j.patrec.2015.10.013. URL http://www.sciencedirect.com/science/article/pii/S0167865515003645.

[23] T. Mueller-Sim, M. Jenkins, J. Abel, and G. Kantor. The robotanist: A ground-based agricultural robot for high-throughput crop phenotyping. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3634–3639, May 2017. doi: 10.1109/ICRA.2017.7989418.

[24] Onur Özyesil, Vladislav Voroninski, Ronen Basri, and Amit Singer. A survey on structure from motion. *CoRR*, abs/1701.08493, 2017. URL http://arxiv.org/abs/1701.08493.

[25] Michael P Pound, Jonathan A Atkinson, Darren M Wells, Tony P Pridmore, and Andrew P French. Deep learning for multi-task plant phenotyping. *bioRxiv*, 2017. doi: 10.1101/204552. URL https://www.biorxiv.org/content/early/2017/10/17/204552.

[26] David Rousseau and Henricus J. Van de Zedde. Counting leaves without "finger-counting" by supervised multiscale frequency analysis of depth images from top view, 01 2015.

[27] Hanno Scharr, Tony P. Pridmore, and Sotirios A. Tsaftaris. Editorial: Computer vision problems in plant phenotyping, cvppp 2017 – introduction to the cvppp 2017 workshop papers. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[28] Kyle Simek and Kobus Barnard. Gaussian process shape models for bayesian segmentation of plant leaves. In H. Scharr S. A. Tsaftaris and T. Pridmore, editors, *Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP)*, pages 4.1–4.11. BMVA Press, September 2015. ISBN 1-901725-55-3. doi: 10.5244/C.29.CVPPP.4. URL https://dx.doi.org/10.5244/C.29.CVPPP.4.

[29] Jack Sklansky. Finding the convex hull of a simple polygon. *Pattern Recogn. Lett.*, 1(2):79–83, December 1982. ISSN 0167-8655. doi: 10.1016/0167-8655(82)90016-2. URL http://dx.doi.org/10.1016/0167-8655(82)90016-2.

[30] Siddharth Srivastava, Swati Bhugra, Brejesh Lall, and Santanu Chaudhury. Drought stress classification using 3d plant models. *CoRR*, abs/1709.09496, 2017. URL http://arxiv.org/abs/1709.09496.

[31] Jordan Ubbens, Mikolaj Cieslak, Przemyslaw Prusinkiewicz, and Ian Stavness. The use of plant models in deep learning: an application to leaf counting in rosette plants. *Plant Methods*, 14(1):6, 2018. ISSN 1746-4811. doi: 10.1186/s13007-018-0273-z. URL https://doi.org/10.1186/s13007-018-0273-z.

[32] Kesheng Wu, Ekow Otoo, and Kenji Suzuki. Optimizing two-pass connected-component labeling algorithms. *Pattern Anal. Appl.*, 12(2):117–135, February 2009. ISSN 1433-7541. doi: 10.1007/s10044-008-0109-y. URL http://dx.doi.org/10.1007/s10044-008-0109-y.