Height Estimation of Biomass Sorghum in the Field Using LiDAR

Matthew Waliman and Avideh Zakhor; University of California, Berkeley; Berkeley, CA

Abstract

Efficient plant phenotyping methods are necessary to accelerate the development of high yield biofuel crops. Manual measurement of plant phenotypes, such as height is inefficient, labor intensive and error prone. We present a robust, LiDAR based approach to estimate the height of biomass sorghum plants. A vertically oriented laser rangefinder onboard an agricultural robot captures LiDAR scans of the environment as the robot traverses between crop rows. These LiDAR scans are used to generate height contours for a single row of plants coorsponding to a given genetic strain. We apply ground segmentation, iterative peak detection and peak filtering to estimate the average height of each row. Our LiDAR based approach is capable of estimating height at all stages of the growing period, from emergence e.g. 10 cm through canopy closure e.g. 4 m. Our algorithm has been extensively validated by several ground truthing campaigns on biomass sorghum. These measurements encompass typical methods employed by breeders as well as higher accuracy methods of measurement. We are able to achieve an absolute height estimation error of 8.46% ground truthed via "by-eye" method over 2842 plots, an absolute height estimation error of 5.65% ground truthed at high granularity by agronomists over 12 plots, and an absolute height estimation error of 7.2% when ground truthed by multiple agronomists over 12 plots.

1. Introduction

Understanding the relationship between genotypes and phenotypes of plants is essential to the optimization of the biofuel production pipeline. By accurately characterizing physiological traits of plants, it is possible to determine connections between plant gene sequences and biomass yield. Specifically, sorghum has been demonstrated to be a suitable source of fuel in practice [1, 2]. To determine the genotype-phenotype map for sorghum, rapid phenotyping methods are necessary for efficient data collection. Currently, plant phenotyping is done by hand, presenting a bottleneck in the growing pipeline [3]. Manual collection of physiological traits is labor intensive, time consuming, inaccurate and does not provide nearly enough data to supplement the genotypes available. Therefore, it is essential to develop systems that automatically, accurately and efficiently phenotype in situ plants.

Plant height is of particular to geneticists and breeders as it is a good indicator of biomass yield [4]. We propose a practical and robust method that estimates biomass sorghum height at all stages of the growing period. Specifically, we estimate the height of plant stalks using data collected in field conditions measured by a scanning LiDAR mounted onto a mobile robotic platform, which traverses through rows of densely positioned plants. We propose two algorithms, a peak based method for determining plant height before canopy closure, and a percentile based method after canopy closure. We show height estimation with less than 10% average error across extensive validation campaigns.

2. Related Work

Various height estimation methods have been developed to take advantage of 3D sensors. These can be divided into two major categories: passive and active sensing systems. Current passive sensing methods are particularly susceptible to changes in ambient lighting conditions and suffer from low spatial resolution [5]. These sensitivities are of particular importance in the context of a crop field, where lighting conditions are highly variable and dense foliage offers ample opportunity for full or partial sensor occlusion. LiDAR, an active sensor, is ideal for direct measurement of canopy height and architecture, offering many advantages over passive sensing including: (1) operation despite variable lighting conditions and (2) high spatial resolution offering greater data reliability in the face of occlusions [5, 6].

Existing methods of height estimation via LiDAR primarily use static measuring systems[7, 8, 9]. For example, Phan et. al. [8] mount a 3D laser scanner at a high vantage point overlooking a field and estimate plant height from the 3D point cloud. Friedli et. al. [9] also extract canopy height using 3D laser scanners in fixed locations throughout the field.

We focus on mobile measuring systems in order to leverage the mobility and flexibility of agricultural robot systems with the ability to traverse between crop rows. Previous methods of mobile height estimation make use of a top down approach that employ stereo cameras mounted on a mast on the robot [10, 11]. In addition to being susceptible to issues endemic to cameras such as occlusion and a narrow field of view [11], a top down approach requires that the sensor be mounted on a mast above the canopy, severely limiting the range of plant heights that can be measured and reducing the overall mobility of the agricultural robot.

In this work, we develop a bottom-up LiDAR based approach that overcomes the mast based issues of previous height estimation setups, eliminating the need for readjustment throughout the growing season and allowing the robotic system a greater degree of mobility between dense crop rows.

The rest of the paper is organized as follows: Section 3 describes our experimental setup, in Section 4, we outline our proposed approach; Section 5 includes experimental results. Finally in Section 6, we discuss implications of our findings and future work.

3. Setup

In the 2018 growing season, surveys were carried out on sorghum field at the University of Illinois Urbana Champaign and an University of California Davis. Four field campaigns were carried out between the 8th of June and 10th of October, covering the entire growing season.

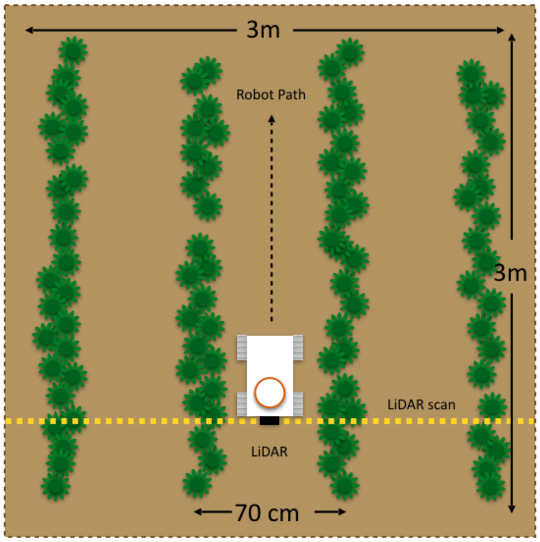

Figure 1: A top down view of a single plot.

Each field is subdivided into plots measuring 3 m by 3 m shown in Figure 1. The Illinois field contained 864 plots while the Davis field contained 250. Each sorghum plot contains either a unique genetic strain or a microbial treatment. Since geneticists are primarily interested in comparing statistical parameters of phenotypes across plots, our goal is to estimate the average stem height for each plot

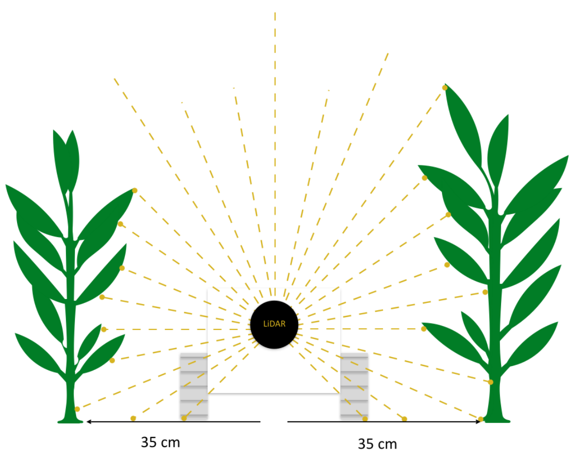

Figure 2: Back view of the sensor set up on the agricultural robot.

Each plot consists of four rows of sorghum, planted by precision planter 70 cm apart from each other. To collect data, a vertical facing Hokuyo UST-10LX is mounted onto mobile robotic platform as in Figure 2, traversing the center row of each plot.

4. Proposed Approach

In this section, we propose two height estimation algorithms: Peak and Percentile methods. The former is used in the early growth season and the latter when the canopy closes.

4.1 Peak Method

As shown in Figure 3, our proposed peak based algorithm consists of four main stages: height contour generation, ground segmentation, adaptive peak detection, and peak filtering. The first stage calculates height for each scan to generate height contours for each plot. Second, the ground segmentation stage applies RANSAC to the resultant height contour to separate points corresponding to the ground from plants. Plant regions are combined into a single contiguous planted region, analogous to the planted region determined by precision planter. Third, adaptive peak detection leverages knowledge of the number of plants planted in each row by precision planter to detect the appropriate number of peaks in a height contour. These peaks are determined by a sliding window and over the planted region. Lastly, peaks are filtered to remove outliers. In what follows, we will describe each step of the algorithm.

4.1.1. Height Contour Generation

Our initial goal is to capture the silhouette of a row of plants, creating a 2D representation of plant heights throughout a row, as show in Figure 4. We refer to this line as a 'height contour'.

We refer to a single rotation of the LiDAR as a single scan as seen in Figure 5. As the robot traverses a row of plants, it collects scans at regular intervals. We limit the range of the scan to 70 cm on either side of the robot in order to restrict the field of view to center rows as shown in Figure 1. In line with manual measurement methods, breeders only measure center rows as they are less affected by fringe effects. To estimate a height contour, we find the maximum y-value for each scan and concatenate these maximum heights together over the entire run of the data collection. An example of the resultant height contour is shown in Figure 6.

4.1.2. Ground Segmentation

Next, we segment the height contour into ground points and planted points. To determine the ground points in the height contour, we run RANSAC on the bottom 10% of all points of the height contour which fits a horizontal line to the ground points. The fitted line and segmented height contour can be seen in Figure 7.

It is necessary to define a single contiguous planted region for each height contour due to gaps in planting and noise from other objects such as humans or equipment in the field caught in view of the LiDAR. For each height contour, we define contiguous regions of more than ten LiDAR return points which were not found to be part of the ground as a planted segments. 'Planted segments' smaller than ten points in width are likely to be weeds or other objects in the field. For each row we find the first planted region and the last planted region and concatenate these together to form a single, contiguous planted region as seen in Figure 8 to run peak detection, ignoring the rest of the height contour.

4.1.3. Iterative Peak Detection

To accurately average the heights of the plants, we distinguish individual plant maximum heights by performing peak detection on the height contour. We define a peak to be any point that is greater than a window of k of its neighbors. k is determined iteratively by the number of peaks found in a single row. The rows are planted by a precision planter that plants anywhere from 25-40 plants per row, so we iteratively run peak detection and adjust

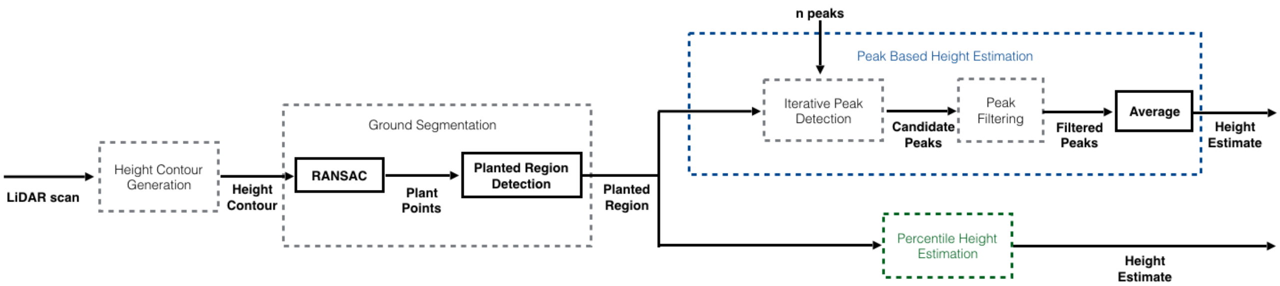

Figure 3: Diagram of proposed pipeline.

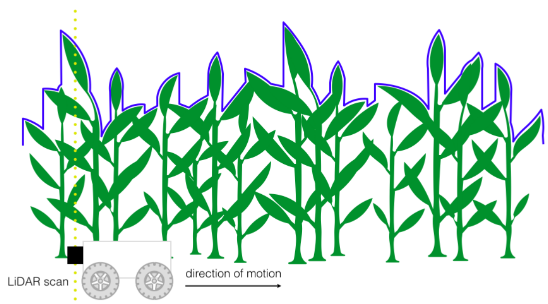

Figure 4: Agricultural robot travels across a row, collecting vertical LiDAR scans. The blue line represents the height contour generated by concatenating these scans in the direction of motion of the robot.

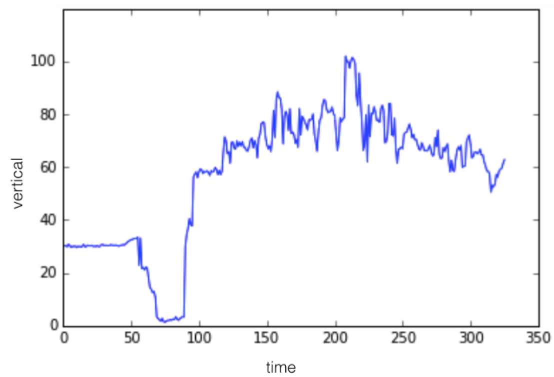

Figure 6: Height contour for a single row.

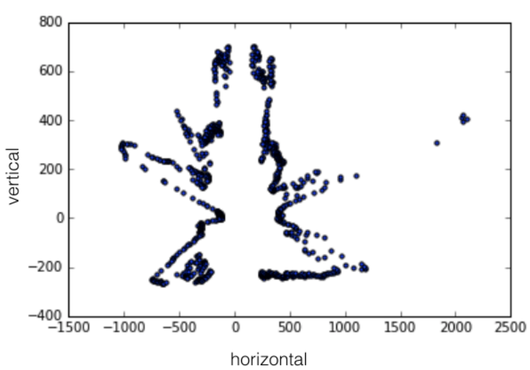

Figure 5: A single LiDAR scan.

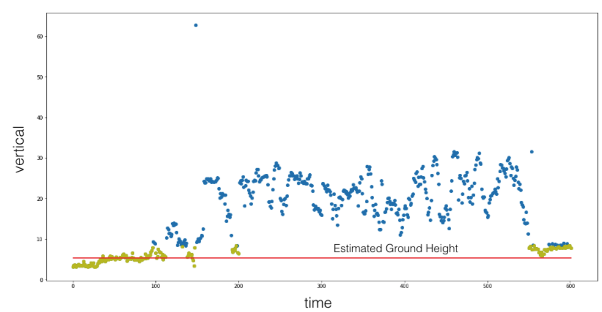

Figure 7: Ground segmentation of height contour. Blue points correspond to planted points, yellow points are determined to be ground points. The red line shows the estimated ground height found by averaging ground point heights.

k until we have found between 25-40 peaks in a single contiguous planted region as shown in Figure 9.

4.1.4. Peak Filtering

There may be weeds, sickly plants or other sources of noise in the field that can affect the detection at the terminations of the detected planted region. These are reflected in the final shape of the height contour and by erroneous peaks at the edges of single planted region. We leverage the fact that the robot travels between two rows and so both left and right height contours for a single plot are calculated at the same time. If either the left or right row suffers from one of these types of noise, we are able to account

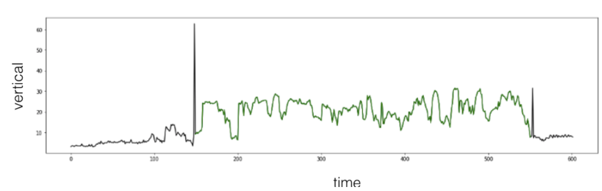

Figure 8: Portion of the height contour in green denotes the single planted region found.

for it by discarding all the peaks found outside the bounds of the

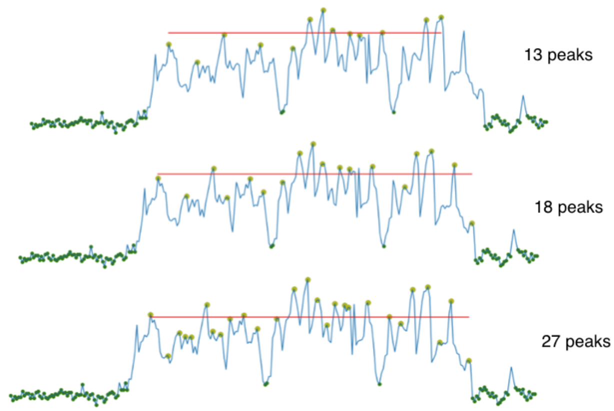

Figure 9: Number of peaks detected increases as k is increased.

other row. In Figure 10 peaks that have been removed due to inconsistencies between left and right rows are shown in blue. We found this to be a robust method to remove erroneous peaks at the edges of the generated height contour.

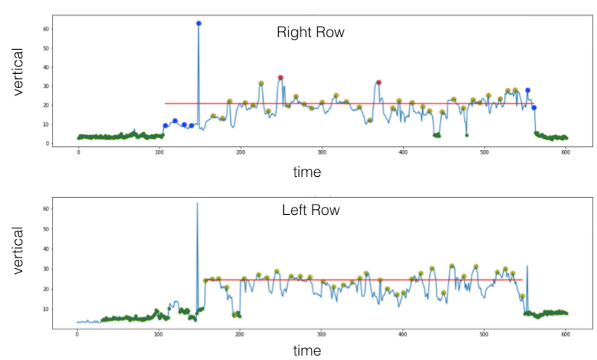

Figure 10: Right and left contours for a single plot. Green points are part of the ground region and are not eligible for peak detection. Blue points in the right contour are outside of the bounds the left row and are removed. Red points are filtered because they are too far from the mean. The red line shows the estimated height by the peak method.

Finally we perform simple outlier filtering of peaks found outside of two standard deviations of the peak average, either due to reflections or objects moving in view of the LiDAR. These filtered peaks are shown by the red points in Figure 10. We average together the the heights of the remaining peaks to determine the estimated height for the plot. The estimated height is represented in Figure 10 by the red horizontal line.

4.2. Percentile Method

Our peak based method assumes that peaks in the height contour correspond to individual plant stems in the row. While this may be generally true before canopy closure, after the canopy closes, individual peaks are not representative of individual stems. In these cases the performance of our peak based method suffers resulting in a significant underestimation of plant heights.

We propose a percentile method for which the first two stages of the algorithm, height contour generation and ground segmentation are identical. However, rather than detecting and filtering peaks, we simply take a top percentile of the 'planted' points. These planted points are points remaining in the height contour after the ground has been segmented out. The 75th percentile is determined empirically so as to result in a less than 10% error across all datasets. It is important to note that the optimum percentile was not the same across all datasets.

5. Results

We present three experiments to evaluate our algorithm in several ways. First, we show that our algorithm outperforms conventional hand measurement methods and generalizes to all stage of the growing season for plant heights ranging from 10 cm to 4 m. In our second experiment, we determine the accuracy of our method. Finally, we measure the variability in human measurement use to evaluate our performance throughout these experiments.

5.1. Experiment 1: Comprehensive By-Eye Measurement

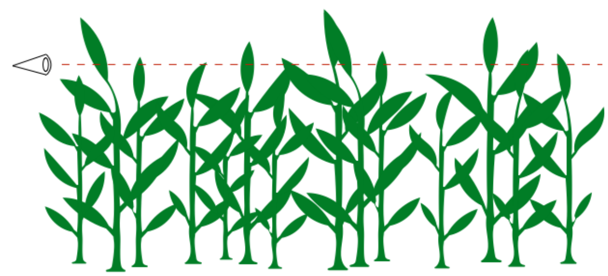

Figure 11: By eye method of height measurement. A single measurement is recorded per plot.

To mimic existing methods of phenotype data collection, our team collected ground truth height data via traditional methods for 864 plots with varying genetic strains over three dates as well as an additional 250 plots at a fourth date. Measurement was performed by the by-eye method currently employed by breeders. This entails placing ones eye to the level of one of the rows and visualizing the line that encompasses the majority of the tops of the plants as shown in Figure 11. The distance from the visual line to the ground is the height estimate for the plot. The plants ranged from 5 cm - 120 cm.

Date	Number of Plots	Height Range (cm)	Average % Error	Average Absolute % Error
6/8	800	0-30	6.6	12.6
6/25	800	50-100	3.2	5.1
10/3	250	80-120	-7.2	8.3
8/3	800	120-400	-9.3	11.7

Table 1: Comprehensive By Eye Measurement: Peak Method.

Table 1 shows the peak based height estimation error as well as the number of plots in each data collection for Experiment 1. As seen, for each data collection we achieve an average % error

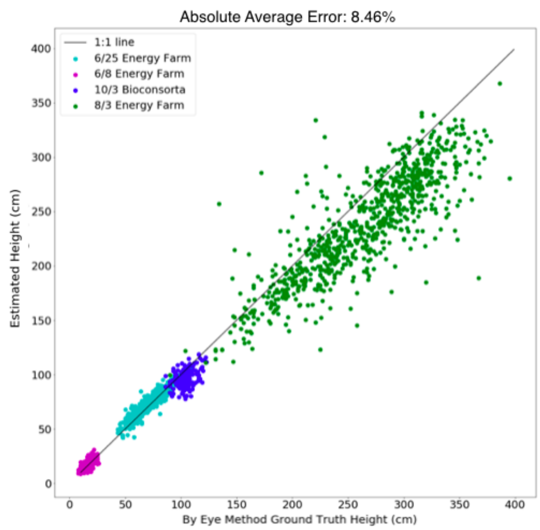

Figure 12: Comprehensive By Eye Campaign: Peak Method. Measured height is on the x-axis and the estimated height is on the y-axis.

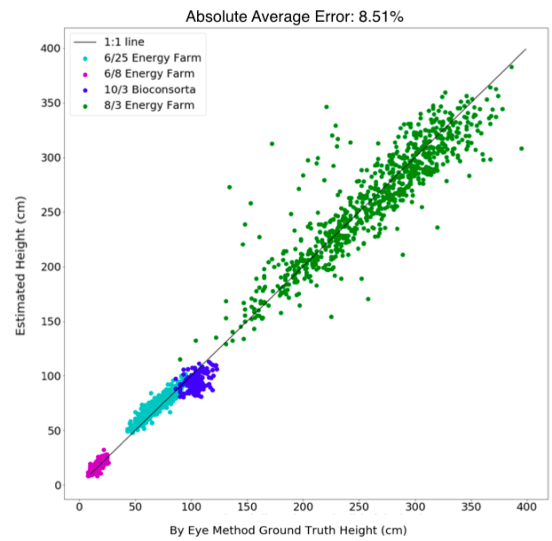

Figure 13: Comprehensive By Eye Campaign: Percentile Method. Measured height is on the x-axis and the estimated height is on the y-axis.

less than 10 %. Figure 12 plots the results of the Experiment 1 on a correlation graph where the x-axis denotes the measured ground truth and the y-axis shows the estimated lidar height, resulting in an absolute average error of 8.46%. The datasets with the greatest average absolute error are the June 8th and August 3rd datasets, representing the two extremes in terms of plant height. The poor performance of the June 3rd dataset is explained by the minute stature of the plants this early in the growing season. Plants at the time of collection ranged in height from 0-30 cm, resulting in tighter margins of error. The poor performance of the the August 3rd dataset is due to canopy closure, where all the plants are greater than 120 cm, and a peak based method of height estimation may not be suitable.

Date	Number of Plots	Height Range (cm)	Average % Error	Average Absolute % Error
6/8	800	0-30	9.1	12.9
6/25	800	50-100	4.7	5.7
10/3	250	80-120	-8.8	9.7
8/3	800	120-400	0.3	6.6

Table 2: Comprehensive By Eye Measurement: Percentile Method.

Table 2 shows the height estimation error for the percentile based method and Figure 13 plots the results of the percentile based method on a correlation graph. We achieve similar performance across all four datasets. The average absolute height estimation error for the percentile method is 8.51% which is slightly higher than that of the peak method with 8.46%. The percentile based method generally overestimates more for plants earlier in the growing season compared to the peak base method. After canopy closure, the percentile based method performs better than the peak based method, underestimating far less, resulting in a 0.3% average error compared to a -9.7% average error via the peak based method.

5.2. Experiment 2: High Resolution Measurement

Figure 14: High resolution method of height measurement. Measurements are taken every 10 cm of the row and averaged together per plot.

To develop more rigorous measurements to test the accuracy of our algorithm, we measured five plots with varying genetic strains over two dates. For the two center rows of each plot, we took measurements every 10 cm for the length of the plot as shown in Figure 14. We averaged these measurements per plot to arrive at the final height measurement for each plot. The plants ranged from 10 cm - 100 cm.

The results of Experiment 2. are plotted via correlation graph in Figure 15 where the x-axis denotes the measured ground truth and the y-axis shows the estimated lidar height. Compared to Experiment 1, we achieve a much lower average absolute error of 5.65%. Table 3 shows the performance of the peak method on each of the individual plants. Again we note that the plots with the poorest performance are ones where the plants were among the

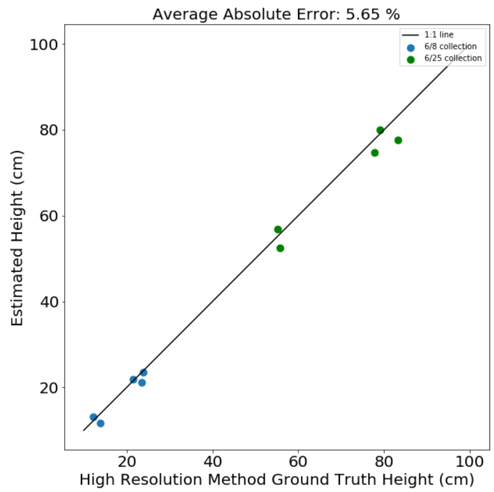

Figure 15: High Resolution Measurement. Measured height is on the x-axis and the estimated height is on the y-axis.

Date	Plot Name	Measured Height(cm)	Estimated Height(cm)	Average % Error
6/8	1	21.4	21.8	1.9
6/8	2	12.2	13.1	7.6
6/8	3	23.8	23.6	-1.1
6/8	4	13.7	11.6	-15.1
6/8	5	23.5	21.2	-9.9
6/25	1	83.2	77.5	-6.8
6/25	2	55.1	56.8	3.1
6/25	3	77.8	74.7	-4.0
6/25	4	55.7	52.4	-5.8
6/25	5	78.9	79.9	-1.2

Table 3: High Resolution Height Measurement: Peak Method.

shortest measured, with differences between ground truth measurement and height estimates of less than 3 cm.

5.3. Experiment 3: Human Variability

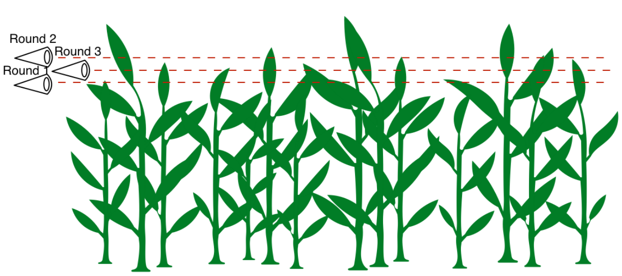

Figure 16: Human variability method of height measurement. Multiple data collectors perform 3 rounds of measurement and their measurements are averaged together per plot.

We measured the variability associated with human measurement of sorghum plants. Specifically, seven researchers independently took measurements of 12 plots of varying genetic strains on the same date in three rounds. Each researcher measured each plot employing the by-eye method three times as shown in Figure 16. The average of all three rounds over all researchers for each plot was used as the "true ground truth" for each plot. The plants ranged from 20 to 210 cm.

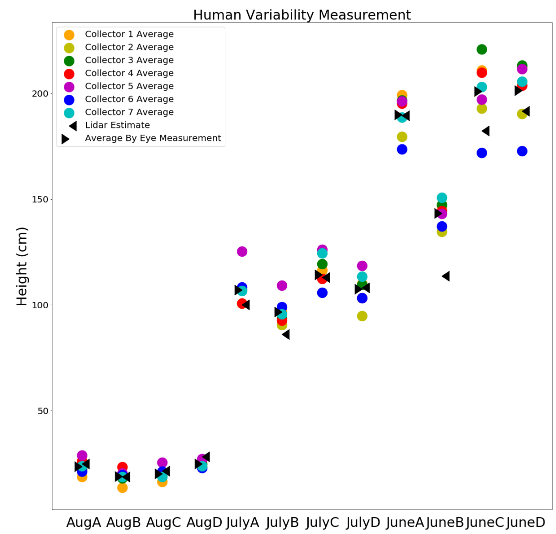

Figure 17: Human Variability Measurement. Plot names are on the x-axis and height is plotted on the y-axis. Each colored dot represents the average height for that collector for that plot. The right and left carets show the LiDAR estimate and average of by-eye measurement across all collectors respectively.

	Average % Variability	Average Absolute % Variability
Collector 1	-4.3	6.9
Collector 2	-4.9	8.8
Collector 3	10.3	11.9
Collector 4	0.9	4.3
Collector 5	0.7	4.6
Collector 6	-5.4	10.0
Collector 7	3.4	8.9
LiDAR	-2.2	7.2

Table 4: Average percent variability for each human collector and LiDAR.

Experiment 3 results are shown in Table 4 and Figure 17. In Table 4 collector variability represents the average difference for each collector from the "true ground truth" across the three trials performed. The LiDAR variability is computed with respect to the "true ground truth" as well. The LiDAR variability is 7.2%, outperforming four out of the seven researchers in terms of absolute average variability. This finding demonstrates that the variability of our LiDAR based method is on par with that of human measurement. It also suggests that a major source of error in our previous results may be due to human variability in the measurement of ground truth and rather than due to inaccuracies in the height estimation algorithms. Furthermore, comparing the increase in performance of the high resolution measurement in Experiment 2 to the performance of the standard by eye measurement used in Experiment 1 suggests that LiDAR based algorithms are closer the actual ground truth height than currently employed by-eye methods of manual measurement.

6. Discussion

In this paper, we introduced two automated algorithms for estimating the average height of a plot of sorghum crops using LiDAR mounted on an agricultural robot. Our experiments verify that our algorithm performs on par with, if not better than currently employed manual methods and that our algorithm generalizes to plants at all stages of the growing season ranging in height from 5 cm to 4 m.

Future work could develop machine learning models that are able to improve the accuracy of our height estimation from height contours, especially at greater plant heights after canopy closure.

7. Acknowledgements

The information, data, or work presented herein was funded in part by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000598. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

[1] S Prasad, Anoop Singh, N Jain, and HC Joshi, Ethanol production from sweet sorghum syrup for utilization as automotive fuel in india, Energy & Fuels, vol. 21, no. 4, pp. 2415 - 2420, 2007.

[2] Mustafa Balat and Havva Balat, Recent trends in global production and utilization of bio-ethanol fuel, Applied energy, vol. 86, no. 11, pp. 2273 - 2282, 2009.

[3] M. Minervini, H. Scharr, and S. A. Tsaftaris, Image analysis: The new bottleneck in plant phenotyping [applications corner], IEEE Signal Processing Magazine, vol. 32, no. 4, pp. 126 - 131, July 2015.

[4] Reynolds, M. P., and Borlaug, N. E. Impacts of breeding on international collaborative wheat improvement. J. Agric. Sci. 144, 3. 2006.

[5] Deery, D., Jimenez-Berni, J., Jones, H., Sirault, X., and Furbank, R. Proximal remote sensing buggies and potential applications for field-based phenotyping. Agronomy 4, 349-379. 2014.

[6] Lin, Y. LiDAR: An important tool for next-generation phenotyping technology of high potential for plant phenomics? Comput. Electron. Agric. 119, 61-73. 2015.

[7] Tilly, Nora and Hoffmeister, Dirk and Schiedung, Henning and Htt, Christoph and Brands, J and Bareth, G. Terrestrial laser scanning for plant height measurement and biomass estimation of maize. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XL-7. pp 181-187. 2014.

[8] Anh Thu Thi Phan, Kazuyoshi Takahashi, Atsushi Rikimaru, and Yasuhiro Higuchi, Method for estimating rice plant height without ground surface detection using laser scanner measurement, Journal of Applied Remote Sensing 10, no. 4 046018 - 046018 2016.

[9] Michael Friedli, Norbert Kirchgessner, Christoph Grieder, Frank Liebisch, Michael Mannale, and Achim Walter, Terrestrial 3D laser scanning to track the increase in canopy height of both monocot and dicot crop species under field conditions, Plant methods 12, no. 1, 9 . 2016.

[10] Jin, Jihui, Gefen Kohavi, Zhi Ji and Avideh Zakhor. Top Down Approach to Height Histogram Estimation of Biomass Sorghum in the Field. Computational Imaging, vol. 17 2018.

[11] Tavor Baharav, Mohini Bariya, and Avideh Zakhor, In Situ Height and Width Estimation of Sorghum Plants from 2.5 d Infrared Images, Electronic Imaging 2017, no. 17, pp. 122-135. 2017.

Author Biography

Matthew Waliman received his B.S. in Electrical Engineering and Computer Science from the University of California Berkeley in 2018. He is currently a research scientist at Signetron Inc, focusing on 3D signal processing. In the past, he worked at Ayar Labs and as a R&D Engineer at UC Berkeley. His areas of interest include: 3D Computer Vision, Graphics, Optics, HCI and Augmented Reality.

Avideh Zakhor is Qualcomm Chair and professor in EECS at U.C. Berkeley. She has won a number of best paper awards, including the IEEE Signal Processing Society, Circuits and Systems Society and IEEE Solid Circuits Society in 2008. She was a Hertz fellow, General Motors scholar, IEEE fellow, received the Presidential Young Investigators (PYI) award in 1990, and was selected as electronic Imaging scientist of the year by SPIE in 2018.