In-Field High-Throughput Phenotyping of Cotton Plant Height Using LiDAR

Shangpeng Sun 1, Changying Li 1,* and Andrew H. Paterson 2

1 College of Engineering, University of Georgia, Athens, GA 30602, USA; shangpeng.sun25@uga.edu
2 College of Agricultural and Environmental Sciences and Franklin College of Arts and Science, University of Georgia, Athens, GA 30602, USA; paterson@uga.edu

* Correspondence: cyli@uga.edu; Tel.: +1-706-542-4696

Academic Editors: Parth Sarathi Roy, Clement Atzberger and Prasad S. Thenkabail
Received: 26 January 2017; Accepted: 13 April 2017; Published: 18 April 2017

Abstract: A LiDAR-based high-throughput phenotyping (HTP) system was developed for cotton plant phenotyping in the field. The HTP system consists of a 2D LiDAR and an RTK-GPS mounted on a high clearance tractor. The LiDAR scanned three rows of cotton plots simultaneously from the top and the RTK-GPS was used to provide the spatial coordinates of the point cloud during data collection. Configuration parameters of the system were optimized to ensure the best data quality. A height profile for each plot was extracted from the dense three dimensional point clouds; then the maximum height and height distribution of each plot were derived. In lab tests, single plants were scanned by LiDAR using 0.5° angular resolution and results showed an $R^2$ value of 1.00 (RMSE = 3.46 mm) in comparison to manual measurements. In field tests using the same angular resolution; the LiDAR-based HTP system achieved average $R^2$ values of 0.98 (RMSE = 65 mm) for cotton plot height estimation; compared to manual measurements. This HTP system is particularly useful for large field application because it provides highly accurate measurements; and the efficiency is greatly improved compared to similar studies using the side view scan.

Keywords: precision agriculture; field robotics; LiDAR; high-throughput phenotyping; crop surface model; plant height

1. Introduction

Geometric characteristics of crops (height, crown size, and volume) have great scientific value for crop breeders and geneticists [1,2]. These characteristics are useful not only for quantitative analysis of genotype–environment interactions, which is essential for increasing crop performance [3–5], but also for improving crop management strategies such as fertilization, irrigation, and optimization of harvesting [6–8]. Several studies have shown interactions between geometric parameters of plants and crop yield as well as biomass [3,5,9–11]. Manual measurements of crop traits, which are still often used in practical phenomic applications, have significant limitations and drawbacks. These methods are time-consuming and labor intensive, which inevitably increases cost [9,12]. Additionally, manually obtained measurements are subject to human error due to fatigue and distractions during data collection.

Several platforms based on remote sensing technologies have been developed for automatic crop phenotyping in greenhouses over the past three decades [13,14], but such controlled conditions are significantly different in many important ways from the outdoor agricultural environment in which the vast majority of crops are produced. Admittedly, there are tremendous challenges to developing field-based high-throughput phenotyping (HTP) systems in the highly variable natural environment, including variable natural lighting, high temperature, uncontrolled rainfall, and occlusion of plant features between adjacent plants within a plot or between plots, to name a
few. While genomics now empowers detailed and cost effective analysis of plant genetic composition, field-based high-throughput phenotyping capabilities currently remain a major bottleneck limiting plant genome research and crop improvement [5,12].

Crop height is an important parameter among geometric characteristics; it has important implications for plant health, crop yield and quality, and total biomass [6,15,16]. Crop height also influences the interactions between plants and solar radiation, water, and nutrients [9]. Crop growth rates—as reflected by changes in height—provide an indicator to assess the efficiency and effectiveness of management strategies related to fertilization, irrigation management, pruning, and harvesting [17,18]. Cotton plant height is a function of the extension of main stem internodes, which is strongly influenced by determinacy and growing environment [19,20]. Consequently, it is of great value to develop field-based systems to estimate crop height effectively and efficiently.

Several approaches based on remote sensing technologies have been published in the scientific literature for crop height measurement. Ultrasonic sensors are the most commonly used sensor to measure crop height in agriculture applications [9,12,21]. One of the main drawbacks of an ultrasonic sensor is that the field of view becomes larger as the distance between the sensor and the object increases—reducing the accuracy of measurements—due to the sensor’s relatively wide angle divergence of ultrasonic waves. Additionally, the ultrasonic sensor is sensitive to temperature variations, which limits its outdoor use [12].

Computer vision based techniques have also been employed for crop height measurement: Carlone et al. [22] presented a plant height and crown radius measurement system which used an RGB camera with an inertial measurement unit (IMU) and a GPS mounted on a tractor to collect data weekly. A 3D model of the scanned crop was obtained by applying SLAM (simultaneous localization and mapping) techniques [23] and multi-view stereo algorithms [24] to process the images and sensor data. The system obtained reasonable estimates for crop height and crown radius.

Busemeyer et al. [25] used 3D time-of-flight (TOF) cameras for field phenotyping research. A 3D TOF camera was mounted above plant height for vertical top view measurements of the canopy, an enclosure was used to provide a stable environment for the sensor, and the reported mean relative error (MRE) ranged from 4.3% to 6.5%. However, one major limitation of image based methods is that data quality can be significantly affected by the variable environment, since shadows and sunlight can result in under or over exposure and limit automatic data processing.

LiDAR (light detection and ranging) is a remote sensing technology used to measure the distance to an object of interest based on the TOF principle. This technique has gained increasing attention in precision agriculture and forestry applications due to its high accuracy, reading speed rates, and versatility. Furthermore, LiDAR’s robust operational capacities in a wide range of light conditions make this technology highly favorable for outdoor use [26]. In addition, a 3D structure model of plants can be obtained with a 2D LiDAR when georeferenced sensors are used, from which various structural characteristics have the potential to be derived. A 2D LiDAR collects two dimensional scans in a measured plane, and the third dimension is obtained by moving the sensor along the perpendicular direction to the scanning plane. Usually a LiDAR system is attached to different platforms for various demands. Aerial LiDAR has been successfully used in forest ecosystem applications such as mapping and canopy structural characteristic retrieving. Leeuwen and Nieuwenhuis [27] presented an overview of how LiDAR remote sensing can be used to derive forest structure attributes such as tree height, leaf area, and branch detection, and addressed approaches for fusing LiDAR data with passive optical techniques. LiDAR remote sensing is expected to play a key role in forest ecosystem, biomass, and biodiversity studies over the coming years. Zhao and Popescu [28] explored the capability of LiDAR to remotely map leaf area index (LAI) in a temperate forest of the southern USA. The correlation of two LAI maps between the LiDAR and the optical remote sensor (the GLOBCARBON moderate resolution satellite LAI product) was up to 0.85 for pines, which implied that LiDAR offered a viable tool for mapping LAI. However, aerial LiDAR was not as effective in agriculture crops phenotyping
activities since it has limited capability to provide high resolution information for crops which are much smaller than trees [29].

Terrestrial LiDAR has the potential to provide denser point clouds over a relatively small area, from which high resolution information could be extracted. Therefore, it has been increasingly used in field phenotyping [26]. Llorens et al. [30] compared the use of ultrasonic sensors and LiDAR to capture canopy characterization in vineyards, and they concluded that data obtained with LiDAR is generally more precise than data obtained with ultrasonic sensors. Rosell Polo et al. [31] designed a tractor-mounted LiDAR to scan several kinds of tree and plantations from the side view. Known objects were placed for reference at known positions during data collection, and an algorithm based on the reference points was used to obtain the 3D structure model. For some selected trees the volume estimation based on the 3D model was accurate—the correlation coefficient was up to 0.976 between LiDAR measurements and manual measurements—but the procedure of 3D model reconstruction was very time-consuming as several steps had to be carried out manually. Auat Cheein et al. [32] compared four real-time canopy volume estimation approaches based on LiDAR data collected using the methodology [31] and 3D model reconstruction procedure. Results showed that the volume could be estimated using partially scanned canopies on either left or right sides. For the worst case, the volume was about 70% of the value estimated by fully scanned canopies. Other methods for measuring tree canopy structure information using LiDAR were discussed by Pforte et al. [33] and Sanz-Cortiella et al. [34]. Shi et al. [35] introduced a corn plant locations and stalk numbers measurement system using a LiDAR with an encoder mounted on a modified golf cart. However, one limitation was that error might be accumulated since plant locations were determined from LiDAR position information recorded with a shaft encoder. Zhang and Grift [16] proposed a LiDAR-based crop height measurement system for Miscanthus giganteus, in which the LiDAR was attached to a tractor and the crop scanned from the side view. In static mode, an average error of 5.08% was achieved; in dynamic mode, an average error of 3.8% was achieved. For dynamic situations, the tractor must be driven at constant velocity since no odometers were used to provide position information of LiDAR scans. However, moving at a constant speed was not easy to implement in the field. 3D LiDAR is being developed and explored for better characterization of plant structures, and some models are gradually becoming commercial available [36] but are still challenging to operate in harsh field environments. A detailed review of the potential of LiDAR in present and future applications was presented by Bhardwaj et al. [37].

In this study, the height measurement of cotton plants was conducted using a LiDAR, but several improvements were made over the previously mentioned studies. Since cotton plants are much smaller than trees and have what some suggest to be the most complex structure of all major field crops [38], more accuracy is required for a height measurement system for cotton than trees and other crops. In addition to height, other plant structure traits such as crown size and leaf area index can be extracted from a 3D map of cotton plants. The quality of the reconstructed 3D map of cotton plants plays a significant role in trait extraction. An RTK-GPS system which can offer 1 cm accuracy was used to provide the precise spatial coordinates at all times. This was useful for reconstruction of the 3D model because precise GPS data remedied the negative effect of inconsistent tractor speed. In addition, the LiDAR was set to scan plants from the top view so multiple rows could be scanned simultaneously, thereby greatly improving the throughput capability. The LiDAR scanning strategy in this study was similar to the method used by Andújar et al. [39], where two rows of poplar trees were scanned at a time. Side view scanning is another commonly used strategy because other traits in addition to crop height could be measured, but only one row can be scanned simultaneously. We also proposed a systematic approach to optimize system configurations to ensure that the tallest parts of plants were scanned. The method used in this study is particularly useful for field application. Custom software for data acquisition was developed using LabVIEW 2014 (National Instruments Inc., Austin, TX, USA), which can be easily extended and customized.
The overall goal of this work was to develop a high-throughput phenotyping system using LiDAR for cotton plant height measurement in the field. Specific objectives were to: (1) design and implement a robust LiDAR data acquisition system (hardware, software, and system configuration methods) for data collection under dynamic field conditions; (2) develop a point cloud data processing pipeline for GPS and LiDAR data fusion to reconstruct a 3D model of cotton plants and extract height information; and (3) evaluate the performance of the proposed system with lab and field validation tests.

2. Materials and Methods

2.1. Data Acquisition System

A tractor was used as the data acquisition platform (Figure 1a), and the schematic of the system is shown in Figure 1b. A LiDAR sensor unit (LMS 511 PRO SR, SICK AG, Waldkrich, Germany) was attached to a sensor bar at the back of the tractor (John Deere 6000 Sprayer, Deere & Company, Moline, IL, USA). The mounting height of the sensor bar could be adjusted by the tractor’s hydraulic lift system from 1124 mm to 1824 mm above the ground. An RTK-GPS unit (Cruizer II, Raven Industries Inc., Sioux Falls, SD, USA) was used to provide the spatial coordinates of the LiDAR at all times. The LiDAR and RTK-GPS main characteristics are summarized in Tables 1 and 2. Data collected by LiDAR was recorded by a rugged laptop computer (S400, Getac Technology Corporation, Taipei, Taiwan) through an Ethernet interface. A dense 3D model of cotton plants was obtained when the tractor was driven along rows in the field. In this study, the maximum height (1824 mm) of the LiDAR was configured for field experiments. The LiDAR simultaneously scanned three rows of plants from above (the top view) in the field (Figure 1c,d).

Table 1. LMS 511 technical data.

<table>
<thead>
<tr>
<th>Features</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating range</td>
<td>0–80 m Systematic error ±25 mm (1 m–10 m) ±35 mm (10 m–20 m) ±50 mm (20 m–30 m)</td>
</tr>
<tr>
<td>Aperture angle</td>
<td>190° (−5°~185°) Statistical error 6 mm (1 m–10 m) 8 mm (10 m–20 m) 14 mm (20 m–30 m)</td>
</tr>
<tr>
<td>Scanning frequency</td>
<td>25/35/50/75/100 Hz Interface Ethernet, RS-232, RS-422, USB, CAN</td>
</tr>
<tr>
<td>Angular resolution</td>
<td>0.167/0.25/0.333/0.5/0.667° Supply voltage 24 V (22 W)</td>
</tr>
<tr>
<td>Wave length</td>
<td>Infrared (905 nm) Temperature range −30 °C to +50 °C</td>
</tr>
</tbody>
</table>
Table 2. Cruizer II RTK GPS technical data.

<table>
<thead>
<tr>
<th>Features</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling frequency</td>
<td>1, 5 and 10 Hz</td>
</tr>
<tr>
<td>RTK accuracy</td>
<td>1 cm</td>
</tr>
<tr>
<td>Differential correction</td>
<td>SBAS (RTK)</td>
</tr>
<tr>
<td>NEMA output</td>
<td>GGA, GLL, GSA, GSV, RMC, VTG, ZDA</td>
</tr>
<tr>
<td>Interface</td>
<td>RS-232, USB, CAN</td>
</tr>
<tr>
<td>Supply voltage</td>
<td>9–16 V DC (600 mA)</td>
</tr>
<tr>
<td>Temperature range</td>
<td>−30 °C to +70 °C</td>
</tr>
<tr>
<td>Memory</td>
<td>FLASH, 256 MB</td>
</tr>
</tbody>
</table>

Custom software for data acquisition was developed using LabVIEW 2014 (National Instruments Inc., Austin, TX, USA). The front panel and flowchart of the block diagram are shown in Figure 2. A finite-state machine (FSM) was used in the event-driven custom software so the configurations can be changed without stopping the program. The FSM is a mathematical abstraction machine that can store one of a finite number of states at any given time and change from one state to another in response to external inputs. Compared to the commercial software SOPAS (SICK Open Portal for Application and Systems Engineering) which is commercially available software for communication, configuration and data logging, the developed software can be easily extended and customized. For example, it is easy to integrate extra sensors into the system or develop new data processing functions in the software.

Figure 2. Custom data acquisition software. (a) Front panel of the custom data acquisition software; (b) Flowchart of block diagram of the custom data acquisition software.
2.2. Configuration of System Parameters

Cotton plants have perhaps the most complex structure of all major field crops. They do not have a homogenous crop surface like rice and maize. Therefore, a dense 3D point should be obtained to ensure that the highest points of plants are scanned by LiDAR. The interspace between two consecutive measured points within each frame, $\Delta D$, should be less than the diameter of the laser beam, $\Phi$ (Figure 3, Equation (1a)). Similarly, the interspace between two consecutive frames, $\Delta L$, should be less than $\Phi$ (Figure 3, Equation (1b)).

\[
\begin{align*}
\Delta D & \leq \Phi, \\
\Delta L & \leq \Phi
\end{align*}
\]

The diameter of the laser beam $\Phi$ increases in size as the distance from LiDAR increases, and the diameter can be determined by Equation (2) [40]:

\[
\Phi = (0.011 \times \text{dis}) + 13 \text{ [mm]},
\]

where $\text{dis}$ is the distance in mm between the LiDAR and the object to be measured. $\Delta D$ is dependent on the angular resolution selected and the distance $\text{dis}$, which is related to the mounting height. $\Delta L$ is a function of the moving speed of LiDAR—which is the velocity of the tractor—and the scanning frequency (frames per second) in Hz.

2.2.1. Angular Resolution

Figure 4 shows a geometric model to determine the angular resolution of LiDAR. A and B are two adjacent measured points. Distances from LiDAR to these two measured points are OA and OB, respectively. Let $\theta$ be the configured angular resolution, the angle $\varphi = \angle O'OB$, and $H$ the mounting height of LiDAR (Figure 4). The laser diameter of measured points A and B can be calculated using Equation (2), and the interspace between the measured points A and B using Equations (3)–(5). $D_2$ and $D_1$ are the distances from the LiDAR to points A and B, respectively, in the horizontal direction.

\[
\begin{align*}
D_1 &= H \tan(\varphi), \\
D_2 &= H \tan(\varphi - \theta), \\
\Delta D &= D_1 - D_2 = H[\tan(\varphi) - \tan(\varphi - \theta)] \\
&= H \left[\tan(\varphi) - \frac{\tan(\varphi) - \tan(\theta)}{1 + \tan(\varphi) \cdot \tan(\theta)}\right], \\
&= H \frac{\tan(\theta)(1 + \tan^2(\theta))}{1 + \tan(\varphi) \cdot \tan(\theta)}
\end{align*}
\]
If a coarse angular resolution was selected (Figure 5a), $\Delta D$ would be greater than $\Phi$ and some points of the objects would be missed. If a proper angular resolution was selected (Figure 5b,c), there would be no gap among consecutive measured points for each frame.

![Geometric model for determining angular resolution of LiDAR.](image)

Figure 4. Geometric model for determining angular resolution of LiDAR.

2.2.2. Mounting Height

As explained in Section 2.2.1, the mounting height plays a significant role for the system since it affects the distance from LiDAR to the measured points. Mounting height is also an important factor influencing the number of rows that can be simultaneously scanned by LiDAR. In order to maximize throughput, LiDAR should simultaneously scan as many rows as possible without missing any plants. Figure 6a shows an example in which one plant in row 2 could not be scanned by LiDAR since it is blocked by its neighboring plant. Figure 6b–d demonstrate the relationship between the mounting height and the number of rows that can be simultaneously scanned. Here, we only considered the situation in which the LiDAR was mounted directly above the middle plant, and in this situation, only odd numbers of rows were scanned by LiDAR. Figure 6b,c show one and three rows being scanned, respectively. For these two situations, the mounting height should be higher than the tallest plants. Figure 6d illustrates a model used to retrieve the relationship between the mounting height and the scanned rows when five rows are scanned. Based on the two similar triangles with red dashed lines, Equation (6) was derived, which can be rearranged to obtain the mounting height as shown in Equation (7).

\[
\frac{\Delta h}{H-h_2} = \frac{d}{(N-1)d} = \frac{2}{N-1}
\]

(6)

\[
H = \frac{(N-1)h_1}{2} + \frac{(3-N)h_2}{2}, \quad N = 5,
\]

(7)

In Equations (6) and (7), $d$ is the interspace between two adjacent rows; $N$ is the number of rows simultaneously scanned; $H$ is the mounting height of LiDAR; $h_1$ and $h_2$ are the height of plants $p_1$ and $p_2$, respectively; and $\Delta h = h_1 - h_2$, the difference between the heights of plants $p_1$ and $p_2$. Here,
the extreme situation is that \( h_1 \) is the height of the tallest plant, and \( h_2 \) equals zero, which means there is no plant at the \( p_2 \) position.

The LiDAR is able to scan a larger range when it is attached at a higher height, but the laser spot size increases as the distance from the LiDAR to the measured point increases (Equation (2)), resulting in a decreased spatial resolution. Therefore, the accuracy will decrease when the LiDAR measures a point at a far distance.

![Diagram showing LiDAR scanning at different heights and the relationship between mounting height and scanned rows.](image)

**Figure 6.** Relationship between mounting height and scanned rows: (a) diagram of one missed plant in Row 2; (b) one row was scanned; (c) three rows were scanned; (d) five rows were scanned.

### 2.2.3. Moving Speed of the Sensor Unit

Similar to angular resolution, the interspace between two consecutive frames \( \Delta L \), which is the sampling resolution along the direction in which the sensor is moving, should be less than the diameter of the laser beam. \( \Delta L \) can be determined by Equation (8):

\[
\Delta L = \frac{v}{f},
\]

where \( v \) is the velocity of LiDAR (which is the tractor’s speed), and \( f \) is the sampling frequency of LiDAR in frames per second.

### 2.3. Experimental Setup

#### 2.3.1. Lab Experiment Setup

To evaluate the height measurement approach and validate the parameter configuration methods, a wooden frame was constructed and experiments were conducted in the lab. The LiDAR was mounted on the frame and used to scan cotton plants from above to obtain the top view while a cotton plant was put on a cart and moved in and out of the frame during tests (Figure 7). Five cotton plants were used...
for lab tests (Figure 8): plants 1, 3 and 5 were in leaf and canopy development growth stage, while plants 2 and 4 were in flowering and boll development growth stage. Each plant was scanned ten times for each configured angular resolution; the average of the measurements was used as the LiDAR measured height. Manual measurements were used as a reference to evaluate the system performance. Cotton plant height, the distance between the highest point of the plant and the ground, was measured using a roller tape. The variable of angular resolution was used as an example to demonstrate the impacts of different configuration values on measurement performance. Plants were scanned under 0.33° and 0.5°, respectively. The data for angular resolution of 1° and 3° was obtained by taking every second and sixth data point from the data set for 0.5°, respectively. Angular resolutions of 1° and 3° were applied to demonstrate impacts on measurement performance more clearly. During lab tests, the LiDAR was attached at a fixed height of 1803 mm, while the scanning frequency was configured to be 50 Hz, and the cart was manually pushed slowly. Although no exact speed value was measured, we estimated the average speed ranged between 0.09 m/s and 0.12 m/s, which ensured that there was no gap between two consecutive frames. The mounting height value was planned to be the same as the mounting height of field experiments (1824 mm), but since the wooden frame was built manually, the dimension was not maintained well resulting in 21 mm difference.

![Scheme of height measurement for lab tests.](image)

**Figure 7.** Scheme of height measurement for lab tests. The LiDAR was attached on the frame at a height of 1803 mm. Points A and B are two terminal points of the frame at ground base, and O is the projected point of the LiDAR; let O be the original point, A is the position of –500 mm, and B is 500 mm.

![Plants used in lab tests.](image)

**Figure 8.** Plants used in lab tests.

The system should be able to fully scan cotton plants of various heights within a plausible range. Therefore, the plants used for the lab tests were different heights. The tallest and shortest plants represent the two extreme situations. In lab experiments, the tallest plant was plant 2 with a height of 1295 mm, and a height of 0 mm was used as the shortest plant, which meant that the seed was not geminated. For the lowest height situation, the LiDAR scanned the ground base.
The height of the scanned cotton plant was derived using Equation (9):  
\[ H_p = H_L - H_C - \min(H_S), \]  
where \( H_p \) is the height of the cotton plant; \( H_L \) is the mounting height of the LiDAR (\( H_L = 1803 \) mm); \( H_C \) is the height of the cart (\( H_C = 328 \) mm); and \( \min(H_S) \) is the minimum distance between the LiDAR and the plant (or the maximum plant height), measured by the point cloud.

2.3.2. Field Experiment Setup

Field tests were conducted in Watkinsville, GA, USA (33.86631°N, −83.54592°E) in autumn of 2015 (Figure 9). The maximum wind speed was 16 km/h. At that time, cotton was in mature growth stage. In the field, there were 6 rows, each containing 20 plots. Each plot was sowed with 15 seeds with 152 mm between adjacent seeds. The alley length was 1830 mm and the row space was 914 mm. There were 6 cultivars, which were randomly planted. The start and end point positions for each plot were measured with the RTK-GPS.

![Figure 9](image.png)

**Figure 9.** Design of field experiments and the moving direction of the tractor.

During data collection, the LiDAR mounting height was 1824 mm with an angular resolution of 0.5° and a scanning frequency of 50 Hz. The tractor was driven down each row from row 1 to row 6 and three rows were simultaneously scanned for each of the four middle paths. For the border paths (along rows 1 and 6), data for only two rows were used for further processing. The speed of tractor was between 0.45 and 0.54 m/s. For field tests, the maximum height of each plot was measured manually to provide a reference. Similar to lab experiments, cotton plant height was measured using a roller tape as the distance between the highest point of the plant and the ground.

2.4. Data Processing Algorithm and Performance Evaluation

Figure 10 illustrates the data processing pipeline for the LiDAR data. Four steps were contained in the method:

**Step 1** Read raw data

The LiDAR scanned frames including timestamps, and GPS tags were retrieved from test files by a program developed in MATLAB 2016a. The raw LiDAR data is shown in Figure 10a,b.

**Step 2** Preprocess the point cloud of LiDAR

The raw data along the Y axis was processed using Equation (10):  
\[ H_{pt} = H_m - H_{scan}, \]  
where \( H_m \) is the height of the cart (\( H_m = 328 \) mm); and \( H_{scan} \) is the minimum distance between the LiDAR and the plant (or the maximum plant height), measured by the point cloud.
where $H_m$ is the LiDAR mounting height, $H_{\text{scan}}$ is the detected distance between each scanned point along the $Y$ axis and the LiDAR, and $H_{pt}$ is the real height value of each scanned point. Distance filters were then used to eliminate out-of-range data along the $Y$ and $Z$ axis. Along the $Y$-axis direction, the distance filter threshold was determined by the rows scanned by the LiDAR. For example, if three rows were scanned at a time, the threshold would be 1372 mm which is equal to 1.5 times the length of interspace between two rows. Along the $Z$-axis direction, the threshold is set to be the mounting height of the LiDAR; points below the LiDAR were kept. An example of reconstructed 3D point clouds is shown in Figure 10c. The $Y$ axis was the direction of cross section, the $Z$ axis depicted the plant height, and the $X$-axis denoted the tractor’s direction of movement. The $Y$-axis and $Z$-axis indicate lengths in millimeters, while the $X$-axis indicates the number of scanned frames since GPS data was not merged.

Step 3 Fuse GPS data with LiDAR data

GPS data was used to make the conversion of the unit of the $X$-axis from the frame number to distance. Let $P_{\text{GPS}}$ be the set of collected GPS data, and $F_{\text{LiDAR}}$ be the set of the scanned frames of LiDAR (Equation (11)). The number of GPS points was $N$, and the number of LiDAR frames is $M$. The GPS data and LiDAR data were synchronized using timestamp.

$$
\begin{align*}
P_{\text{GPS}} &= \{p_{\text{GPS}0}, p_{\text{GPS}1}, \ldots, p_{\text{GPS}(N-1)}\} \\
F_{\text{LiDAR}} &= \{f_{\text{LiDAR}0}, f_{\text{LiDAR}1}, \ldots, f_{\text{LiDAR}(M-1)}\}.
\end{align*}
$$

(11)

The distance between two adjacent GPS points denoted by $\Delta P_{\text{GPS}}$ was computed by Equation (12). $f_{\text{LiDAR}}$ and $f_{\text{GPS}}$ were the data acquisition frequency of LiDAR and GPS, respectively. In this study, the data acquisition frequency of GPS was $f_{\text{GPS}} = 5 \text{ Hz}$, and LiDAR scanning frequency was $f_{\text{LiDAR}} = 50 \text{ Hz}$. Therefore, there were 10 scanned frames, each containing 381 points (The aperture angle was 190° with angular resolution 0.5°), between every two adjacent GPS points (Equation (13)). Assume that the tractor was moving at a constant speed during the interval (200 ms) of two adjacent GPS points. The distance of the two adjacent frames within two adjacent GPS points was computed using Equation (14). Therefore, the position of each LiDAR scanned frame was able to be obtained using Equation (15). $D_{\text{offset}}$ was the offset between LiDAR and GPS. In this study, $D_{\text{offset}}$ was fixed during data collection in the field, and the measured point at $0^\circ$ scanning angle was used to depict the frame position. Figure 10d showed the reconstructed 3D model with $X$-axis indicated by millimeter units. The point cloud of individual plot was retrieved from the 3D model based on the field layout information (Figure 9). First, the 3D model was segmented into 20 blocks along the tractor moving direction according to the premeasured start and end points of each row and the length of the plot, and then plots within each block were segmented based on the interspace between two rows.

$$
\begin{align*}
\Delta \vec{P}_{\text{GPS}}(i) &= \vec{p}_{\text{GPS}}(i) - \vec{p}_{\text{GPS}}(i - 1), & i &= 1, 2, \ldots, N - 1 \\
\Delta \vec{P}_{\text{GPS}}(0) &= \vec{0}, & i &= 0
\end{align*}
$$

(12)

$$
\alpha = \frac{f_{\text{LiDAR}}}{f_{\text{GPS}}},
$$

(13)

$$
\vec{d}_i = \frac{\Delta \vec{P}_{\text{GPS}}}{\alpha}, & i &= 0, 1, 2, \ldots, N - 1,
$$

(14)

$$
f_{\text{LiDAR}}(k) = f_{\text{LiDAR}}(i\alpha + j) = p_{\text{GPS}}(i) + j \times \vec{d}_i + D_{\text{offset}}, & i &= 0, 1, 2, \ldots, N - 1, & j &= 0, 1, 2, \ldots, 9, & k &= 0, 1, \ldots, M - 1
$$

(15)
Step 4: Extract height trait

A canopy height profile (CHP) of one plot (Figure 10e) would be derived by calculating the maximum height of each frame. Based on CHP, the maximum height and histograms of canopy height were extracted (Figure 10f,g). A threshold of 200 mm was set for the histogram to segment the plants from the ground.

![Diagram of data processing pipeline for LiDAR point cloud]

**Figure 10.** Data processing pipeline for LiDAR point cloud.

The data processing algorithm was developed and implemented in software MATLAB 2016a (The MathWorks Inc., Natick, MA, USA) on a laptop that was used for data collection. The laptop used an Intel i5-4210M CPU 2.60 GHz with 8G RAM, running an operating system of Windows 8.1 Pro.

An accuracy assessment of the system was performed to evaluate the performance of the designed system based on the comparison of LiDAR height measurement and manual measurements. The mean...
error and standard deviation were computed as well as \( R^2 \) value and RMSE, and a simple linear regression analysis was conducted. For field experiments, we performed an accuracy assessment of the system using a total of 288 samples. There were 90, 108, and 90 samples for left rows, middle rows, and right rows, respectively. The first and second plots of all six rows did not germinate, so data from 18 plots (plots 3 to 20) in each row were used for processing. An ANOVA test analysis was conducted to investigate the effect of cultivar on height trait. In addition, algorithm efficiency was tested by performing the algorithm for 100 rounds for each tractor’s path. The average running time with standard deviations was calculated for each path.

3. Results

3.1. Results of Lab Tests

Lab experiments showed that the angular resolution must be at least 0.5° in the following two situations: (1) when the plants have not germinated with the ground base level (0 mm) and (2) when the plants fully have developed at a maximum height of 1623 mm. This angular resolution would meet the full coverage requirement: i.e., the diameter of the laser beam (\( \Phi \)) should be greater than the distance between two adjacent laser points (\( \Delta D \)). The height 1623 mm was obtained by combining the height of plant 2 (1295 mm) and the height of the cart (328 mm). As the angular resolution decreased (i.e., from 0.33° to 3°), the diameter of the laser beam did not change noticeably at both heights. At ground level (Figure 11a), \( \Phi \) was greater than at 1623 mm (Figure 11b) given the same angular resolution, because the scanning distance was longer at ground level. In contrast, \( \Delta D \) increased dramatically as the angular resolution decreased. The line of \( \Delta D \) started to intersect with the line of \( \Phi \) at the angular resolution of 0.5° (indicating \( \Delta D > \Phi \)) both at the ground level and the elevated height. The results suggest that if an angular resolution lower than 1° was used, a full scan could not be obtained and some parts of the scanned objects would not be covered by the laser, resulting in significant measurement errors.

![Figure 11](image-url)
Height measurement results (Table 3 and Figure 12) in the lab were in a good agreement with the analysis in Section 2.2. The system achieved high accuracies for angular resolutions of 0.33°, 0.5°, and 1°, while performance decreased when the angular resolution was 3° (Table 3). For the correlation analysis (Figure 12), although the $R^2$ value was the same for four different angular resolutions, the RMSE for 3° was 7.45 mm which was significantly larger than the value for the other three angular resolutions. If the angular resolution is too low, gaps between two laser points may emerge and this may result in poor measurement performance; if the angular resolution is too high, more storage space and computing resources are needed since more points are generated even though these points may not contribute to enhancing measurement accuracy. Therefore, 0.5° angular resolution was chosen for the lab tests.

The lab tests indicated that the system with fine parameter configurations could measure cotton plant height with high accuracy in the lab (controlled environment). These findings provided useful guidelines for cotton plant height measurement in field conditions.

![Figure 12. Correlation analysis for different angular resolutions (angular resolution = $\theta$).](image)

<table>
<thead>
<tr>
<th>Angular Resolution</th>
<th>Manual Measurement (mm)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height (mm)</td>
<td>599 1295 707 1038 512</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Error (%)</td>
<td>−0.14 0.09 −0.68 −0.47 −0.84 −0.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std 1</td>
<td>2.01 1.72 1.03 4.79 1.38 2.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height (mm)</td>
<td>600 1292 702 1035 508</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Error (%)</td>
<td>0.19 −0.20 −0.71 −0.27 −0.85 −0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>3.75 4.21 2.19 3.42 2.83 3.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height (mm)</td>
<td>599 1292 702 1035 506</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Error (%)</td>
<td>−0.01 −0.26 −0.76 −0.31 −1.26 −0.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>3.90 4.42 2.51 3.16 3.91 3.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height (mm)</td>
<td>595 1289 698 1032 501</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Error (%)</td>
<td>−0.67 −0.46 −1.27 −0.58 −2.15 −1.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>3.18 4.29 2.10 5.02 5.09 3.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Standard deviation.
3.2. Results of Field Tests

Since three rows were scanned for field tests, the corresponding scanning range at ground base level was \([-1372, 1372]\) mm in the horizontal direction (Figure 13). The working hypothesis was that the highest points of cotton plants would not be beyond this range. For field tests, there was much variation in plant growth. For example, some seeds may not have germinated in a plot. The angular resolution at ground base level was selected to ensure that the system was able to make a full scan for all situations. When the angular resolution was set at 1° or 3°, the distance between adjacent laser points was greater than the diameter of the laser point in certain parts (for 1°) or in the entire scanning range (3°). When the angular resolution was set at 0.33° or 0.5°, the diameter of the laser point was always greater than the distance between adjacent laser points. Given that both 0.33° and 0.5° produced satisfactory results, the angular resolution was configured to be 0.5° for field tests.

The tractor speed varied between 0.45 and 0.54 m/s (1.0 to 1.2 mph) during field data collection. In order to ensure that the system would make a full scan along the moving direction of the tractor (the distance between two consecutive scanning planes being less than the laser point diameter), the frame rate was set to 50 Hz. Therefore, the resolution along the moving direction of the tractor was between 9 and 10.8 mm/frame. This distance between consecutive frames was less than the smallest laser point diameter. Based on the data and analyses, no scanning gap would occur either along the LiDAR scanning plane or along the moving direction of the tractor.

![Figure 13](image-url). The varying pattern of the laser beam diameter and the distance between adjacent laser points across a 2743 mm scanning length for the field tests. The LiDAR was at the height of 1803 mm.

Three plots (left, middle, and right rows) were scanned by the LiDAR simultaneously (Figure 14a). The 3D plant model reconstructed from the LiDAR data showed that the scanning points were denser for the middle row (directly below the LiDAR) than the left and right rows (further away from the LiDAR) (Figure 14b). It was also observed that the ground plane was not level, with the elevation decreasing from left to right (Figure 14a). The uneven ground level would result in a system error for measurement results.
Figure 14. Example of 3D reconstruction. (a) 3D view of the 3D structure model; (b) Top view of the 3D structure model.

Overall, the average height error was $-0.02\%$ and the standard deviation of the error was 6.84\% (Table 4), which indicated that plant heights measured by the LiDAR system in naturally illuminated fields were in good agreement with ground truth. The results were also in line with previous studies reported by Zhang and Grift [16]. The system achieved better overall performance for the middle rows than for the left and right rows. For the middle row, the $R^2$ value was 0.99, the RMSE was 54 mm, and the mean of the error was $-0.37\%$. The means of the error for left and right rows were 0.58\% and $-2.62\%$, respectively, which indicated that the system overestimated the height for the left rows and underestimated the height for the right rows. Two factors could contribute to this systematic error: locally uneven ground plane and not well levelled measurement system. These factors could prevent the LiDAR from being parallel to the measured plane. In this study, the measurement system was levelled. However, if the ground plane has a greater gradient or the LiDAR scans a wider range, measurement errors would increase. If necessary, a height normalization process should be applied [22], expressing heights with respect to the local ground plane. In the present study, normalization was not necessary, given that the small slope and the short scanning range would not result in significant negative impact on measurement performance. From the error distribution data (Figure 15), we can
see the measurement errors followed similar distributions with 92.41%, 93.75%, and 92.50% of the total measurements within ±10% of the measurement error for left side, middle, and right side plots, respectively.

<table>
<thead>
<tr>
<th>Table 4. Statistic results of field tests.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
</tr>
<tr>
<td>Mean Error 1 (%)</td>
</tr>
<tr>
<td>Std Error 2 (%)</td>
</tr>
</tbody>
</table>

1 The mean error; 2 The standard deviation of error.

Figure 15. Correlation and error analysis between LiDAR and manual measurements. Two vertical dashed lines in the histograms indicate the ±10% error range.

4. Discussion

The designed system has been demonstrated to be capable of using LiDAR for data collection in a high throughput fashion under field conditions. In this study, one row in the field tests was about 90 m long and the average tractor speed was about 0.5 m/s. Therefore, it took roughly 3 min to finish collecting data along one path (three rows). For the 6-row field, the HTP system was able to finish data collection in about 6 min. This relatively high throughput was a major benefit of the HTP system, compared to the side view scan utilized in similar studies [16,31] and manual measurements. This gain of throughput would be particularly useful if the system was applied in larger fields. If the LiDAR was installed at a higher distance above the plants [41], more rows could be scanned, which would result in even higher throughput. In addition, the proposed data post processing method achieved average running time of 3.22 s for each path with the standard deviation 0.06 s. For the 6-row field, the proposed method could extract the maximum height trait for all plots within 10 s. Each frame of LiDAR data occupied 0.71 KB, and thus the data volume was 35.5 KB/s. Therefore, it was possible to integrate the proposed data analysis method with the proposed system for online height measurement. For the typical experiments mentioned in [7], 20,000 plots are required. If one row has 100 plots, there are totally 200 rows. Assume each plot has the same size to our experimental field. It will take about 16 h to scan the whole field and within 20 min to calculate the maximum height for all plots. Assume a person walks 3 km/h and takes 30 s to measure the height of each plot based on our field work experience, it would need about 200 h to finish all measurements. Mounting the LiDAR at greater height could make the LiDAR scan more rows, however, the likelihood of missing plants would increase. Because if the laser beam is larger than the objects, the total amount of light reflected back to
the sensor is not sufficient, the object may not be detected. In addition, faster tractor speed could also improve data collection efficiency, but there was a maximum speed limitation for the tractor since high speed may lead to gaps between two connective frames. This needs to be further explored.

The study has shown that the method to optimize system configurations ensured that the system made full scans without gaps during field data collection. A high quality 3D plant map, which plays a significant role in cotton plant height measurement and other structural traits extraction [22], was reconstructed with the help of an RTK-GPS. Although some outward-facing plant branches in the left and right rows were beyond the scanning range of the LiDAR (Figure 14), our system settings ensured that the laser pulses were able to scan most parts of the plants including the highest points. These outlier branches would not result in large errors of height measurement. Crommelinck and Höfle [42] stated that for crops with homogenous surfaces such as maize, crop height could be measured with low point density. This study has demonstrated that it is necessary to scan with high point density without gaps due to the complex plant structure of cotton plants [38]. In addition, the dense point cloud could also be helpful for plant volume estimation using the partial scanned plants [32].

The LiDAR measurement results were less stable under field than lab conditions, which was mainly attributed to the uncontrolled environment in the field. Wind and the tractor’s vibration and movement during data collection could make plants move, leading to inaccurate measurements [42]. In some unusual cases, errors were around 40%, which could be caused by movements created by tractor or wind during data collection. Weeds were also a major source of noise for LiDAR data, especially at the early growth stage. These factors could be better controlled in the future by weeding before data collection, collecting data on calm days, and mounting the LiDAR in front of the tractor. In addition, more advanced filtering algorithms [43] should be applied to reduce noise in the raw data.

The height trait could be used to help plant breeders’ select cotton genotypes. For example, plant height is used as an indicator to identify early maturing cotton cultivars [20] that can avoid yield losses due to diseases and insect-pest complexes, and increase economic return by reducing input cost. In this study, histograms of the height distribution showed differences among the six cotton genotypes (Figure 16). An ANOVA test showed that the difference was significant ($p$-value < 0.05). In addition, the height trait also has the potential to predict crop yield and leaf area index [44].

![Figure 16. Height histograms for six cotton cultivars. Cultivars with the same letter in the title are not significantly different.](image-url)
5. Conclusions

The high-throughput phenotyping system developed in this study accurately measured cotton plant height at the plot level under natural illumination in the field. This is advantageous compared to image based methods that require controlled lighting conditions. In addition, the LiDAR in this system can scan three rows at a time with the current mounting height. More rows could be scanned simultaneously by raising the mounting height of the LiDAR, resulting in even higher throughput. This could be particularly useful for large fields. The system also can be used for height measurement of other crops such as soybean, wheat, and rice. Future work will be focused on measuring other traits—such as crown size, plant volume, and leaf area index—from the reconstructed 3D point cloud model.

Acknowledgments: This study was funded jointly by the Agricultural Sensing and Robotics Initiative of the College of Engineering, and the College of Agricultural and Environmental Sciences of the University of Georgia. The project was also partially supported by the National Robotics Initiative grant (NIFA grant No. 2017-67021-25928). The authors gratefully thank Jon S. Robertson, Rui Xu and Yu Jiang for their assistance in planting, field management, and data collection.

Author Contributions: S.S., C.L., and A.H.P. conceived the idea and designed the experiments. S.S. collected and analyzed data. S.S., C.L., and A.H.P. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References


41. Tilly, N.; Hoffmeister, D.; Cao, Q.; Lenz-Wiedemann, V.; Miao, Y.; Bareth, G. Transferability of models for estimating paddy rice biomass from spatial plant height data. *Agriculture* 2015, 5, 538–560. [CrossRef]


© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).