Estimation of cotton structural parameters through object detection using deep learning models

Piyush Pandey

Introduction

Cotton is the most important source of natural fiber in the world with an important impact in the economies of several countries. With the rapid increase in world population, cotton production has to be significantly increased to meet the increasing demands (Zaidi, Mansoor et al. 2018). One of the proven methods for increasing crop yield without increasing cultivated land is the selection and breeding high yielding varieties. The breeding process involves the selection of varieties based on their productivity and resistance to stress as displayed in the field environment. High throughput phenotyping methods are aimed at increasing the throughput of the phenotyping process which has been acknowledged to be the bottleneck in the translation of genetic knowledge into useful production in the field (Furbank and Tester 2011).

High throughput phenotyping methods use imaging techniques to quantify the traits of plants, which brings about a significant improvement in efficiency and accuracy when compared to the manual measurement of traits. This is the result of the use of automated imaging systems for image acquisition, and image processing algorithms for the processing of the acquired image data. An increased number of phenotyping studies using a variety of imaging techniques for the quantification of structural parameters, chemical constituents, and physiological processes in plants have been conducted in recent years (Li, Zhang et al. 2014, Hawkesford and Lorence 2017).

Since image processing and computer vision are crucial steps in the high throughput phenotyping pipelines, the technology used in phenotyping studies closely follows the development of techniques and algorithms in these fields. One of the recent developments in computer vision is the successful use of machine learning models using deep convolutional neural networks for the classification and localization of real world objects in images. This development has been successfully utilized in plant phenotyping studies in particular (Pound, Atkinson et al. 2017), and in agricultural research in general (Kamilaris, Prenafeta-Boldú et al. 2018).

In case of cotton phenotyping, some of the structural traits of importance that are amenable to high throughput phenotyping techniques include plant height, flower counts, boll counts, and internodal distances (Sun, Li et al. 2017, Thompson, Thorp et al. 2019), which are traits that provide information about health, growth status, and ultimately, the yield that can be acquired from a plant.

This study explores the use of deep learning methods for the detection of key plant organs in cotton, followed by further processing of the acquired results to derive semantic information about the cotton plant. As a preliminary study, the detection of cotton bolls and main stalk nodes is studied, followed by the use of this minimal information to derive detailed information about the plant structure. The parameters that we attempt to derive include boll production per node, internodal distances, and branch angles.

A pre-trained region-based convolutional neural network (CNN), Faster R-CNN is used in this study for the detection of cotton parts (Ren, He et al. 2015). In a region-based CNN, the region proposal network generates region proposals and the convolutional network classifies the proposed regions into labels, thus detecting and localizing objects in the image.

The acquisition of detailed structural information about a plant with complex canopy structure such as that of a cotton plant requires the use of three-dimensional imaging technology. While deep learning models with two-dimensional images have been used for the counting of bolls as estimates of yield before (Li, Cao et al. 2017, Fue, Porter et al. 2018), the accurate estimation of bolls per plant or bolls per plot using these techniques faces the challenge of occlusion and of keeping track of repeated counts. The three-dimensional imaging techniques, designed to overcome these challenges, come with their own set of disadvantages in the form of excessive requirement of time and resources both for acquisition and processing of data. The ability of using two-dimensional images to acquire the distribution of plant parts in three-dimensional space is studied in the current study. In order to overcome the limitations of two-dimensional RGB images alone, a test is also conducted using a Microsoft Kinect v2, which consists of an RGB and depth cameras in the same device. Depth images have been previously successfully used in phenotyping projects for cotton as well as for other crops (Jin and Tang 2009, Jiang, Li et al. 2016).

We find that the use of RGB images alone can provide us valuable information about the structural parameters of a plant through the detection of key plant parts using deep learning models. Some limitations of the RGB images, such as the loss of depth information of the pixels is overcome by RGB-depth images acquired from a few angles around the plant. This approach is proposed as a faster alternative to complete three-dimensional imaging of plants, where two-dimensional images with depth values are processed using deep learning and traditional computer vision methods to acquire a detailed semantic structural information about the plant.

Methods

Data acquisition

Images of cotton plants were acquired using a Fujifilm X-A10 camera (Fujifilm Holdings Corporation, Tokyo, Japan) from a cotton field in Watkinsville, GA, USA (33.86631°N, −83.54592°E) in the autumn of 2018. The images were acquired with a focus on a single plant such that the structure of the plant could be observed from the side view images. Four side view images were acquired for each plant from angles that were 90 degrees apart from one another. A top view image was also acquired for each plant.

Several cotton plants were uprooted and placed in pots that were then transferred to the laboratory in Athens, GA, USA. For this preliminary analysis, one plant was selected and placed on a turntable, and RGB images were taken using a Fujifilm X-A10 camera from four angles 90 degrees apart from one another. The imaging was conducted against a consistent blank background to simplify subsequent segmentation and analysis. Similarly, four RGB images and four depth images at angles 90 degrees apart were collected using a Microsoft Kinect v2 (Microsoft Corporation, WA, USA). Matlab Image Acquisition Toolbox (Mathworks, MA, USA) was used for collection of images using Kinect v2. The RGB (1920x1080 pixels) and depth (512x424 pixels) images were taken with minimal interval between the triggers to ensure that the plant or the sensor position did not alter significantly. This was important to ensure the proper alignment of depth and RGB images for the registration of depth values and RGB pixels.

Camera calibration for RGB-depth projection

The Kinect v2 sensors were calibrated for deriving the intrinsic and extrinsic parameters based on Zhang's method using Matlab for acquisition and processing (Zhang and intelligence 2000).

This method involves acquiring multiple images of a planar calibration pattern of known dimensions such that the points on the plane can be easily detected or manually selected for the estimation of projective geometry and camera transformation matrices.

Object detection model

A faster RCNN model was trained using 90 images collected in the field, which included images from side views as well as from the top view as described before. Before being used for training, the images were cropped so that the cotton plant in view would occupy most of the image pixels. Bolls visible in the background but belonging to the plants not on the focus were not labelled as bolls. The model for node detection was also trained using 90 images, which included images collected in the field as well as images that were collected in the laboratory using the uprooted images mentioned before. This was done to increase the visibility of main stalk nodes in the training dataset since the imaging angle could be well controlled in the laboratory. Only the main stalk nodes were labelled, where a node was defined as the point where any branch meets the main stalk. Labeling was done with the annotation tool LabelImg.

In case of object detection applications using deep learning networks, the use of pretrained models has been found to be effective, especially in cases where a limited number of images can then be used to fine-tune the model for the identification of the desired objects. A faster R-CNN model with inception resnet v2, pretrained on the Microsoft COCO dataset (Lin, Maire et al. 2014) was was trained separately for the node and boll detection models using the TensorFlow implementation of Faster-RCNN (Huang, Rathod et al. 2017). Data augmentation was performed by flipping the images horizontally and vertically, and by adjustment to saturation, brightness, and contrast. Adam optimization algorithm was used with the default configuration parameters for updating the network weights, and a constant learning rate of 0.9×10^{-4} was used. In order to prevent overfitting, L2 regularization with a weight of 0.001 was applied to the training process. The batch size was fixed to 2

When using the models for detection of plant parts, a score threshold of 0.5 was used to determine if a proposed bounding box would be considered an actual detection of the plant part. In case of the boll detection model, some large objects were detected by the model to be bolls; a threshold based on the standard deviation of the bounding box areas from the mean area was used to remove these large bounding boxes from the list of detected bolls. In case of node detection, the model sometimes ended up detecting nodes not on the main stalk as the main stalk nodes. This is to be expected since the main stalk nodes and the nodes on the branches of a cotton plant are visually indistinguishable, and the only way to tell them apart is to know the location of the main stalk. Here, the non-main stalk nodes were removed by using a method similar to the one used with the bounding boxes from the boll detection network. Assuming that the main stalk nodes will lie sequentially on the image, we can assume that the branch nodes can be observed as outliers in the overall distribution. This assumption was used in this experiment and was found to give satisfactory results.

Boll counts and structural parameters from RGB images

The trained models for boll and node detection were used with these images. The plant was placed on a turntable and images were taken from four sides, 90 degrees apart. Two images taken at 180 degrees apart were used for the preliminary analysis discussed here.

The cotton plant pixels were segmented from the background by first converting the image to the L*a*b color space, where empirically determined threshold values were used to create a segmentation mask. The segmented binary image was then skeletonized, so that the lines representing the branches were one pixel thick. Additionally, the base of the plants was marked with unique colored markers for identifying the root base, or the ground level. The distance between the two colored markers was measured in order to estimate a distance per pixel value on each image.

Based on the detected nodes and the distance per pixel value estimated with the help of the colored markers, node distances were estimated and compared with the ground truth data acquired manually.

Branch angles in the two-dimensional plane visible in the RGB images were estimated by deriving the angles made by the skeletons of the main stalk and the branch skeleton. Each boll detected on an image was assigned to a main stalk node detected based on the shortest distance through the skeleton pixels. In this process, we start at the mid point of a bounding box representing a cotton boll, and we move one pixel at a time through the skeleton and find the shortest paths to each node that has been detected on the image. Finally, we consider that the node with the shortest distance to this particular boll is the node where the branch producing this particular boll arises from. Figure 1 has a visual description of this image analysis procedure. On the skeleton image showing the bounding boxes corresponding to detected bolls (red boxes) and main stalk nodes (blue boxes), the main stalk is shown in green, and the paths from the bolls to the main stalk and then to the nearest main stalk node are shown in red. On the right panel, the estimated angles made by the branches originating at each detected node are shown.

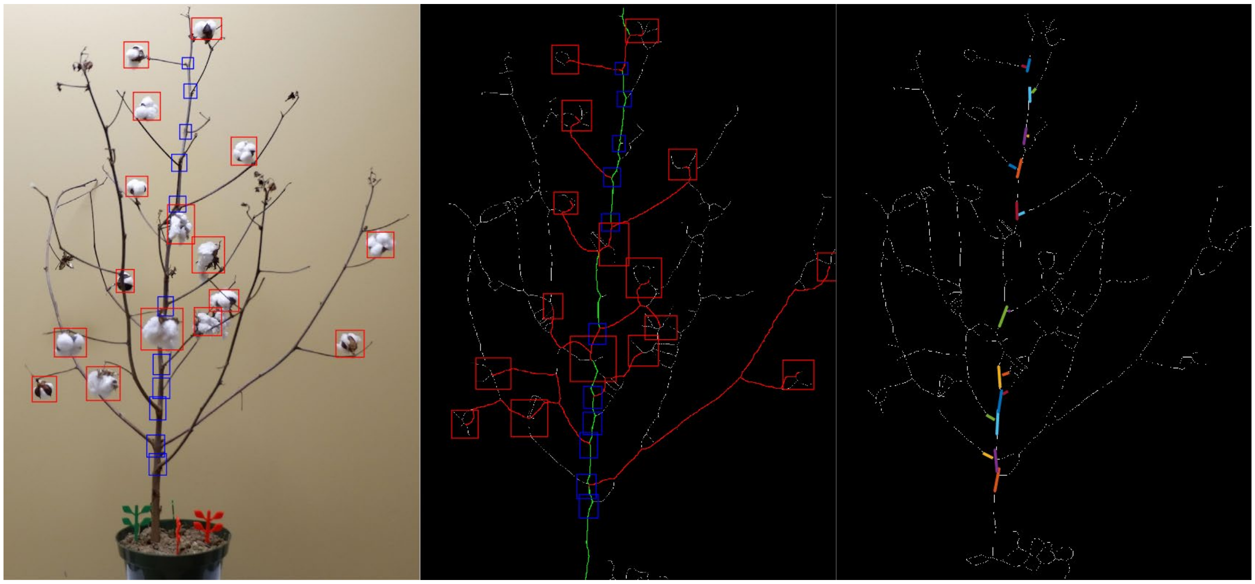

Figure 1 The assignment of bolls to different nodes is based on the estimation of distances from the bolls to the nodes along the skeleton image (center panel), the angle between the branch and the main stalk on each node is estimated assuming a planar geometry for the plant (right panel)

Elimination of repeated counts in multiple views

The problem of repeated counting of the same boll in case of multiple images of the same plant was also studied during this experiment. To do this, the coordinates of the bolls detected in view 1 and view2 were converted to distance units so that the image resolution would not be an issue. To do this, the measured distance between the visible colorful markers placed at the base of the plants was used. As shown in figure 1, the markers are brightly colored 3D-printed objects and could be readily segmented using color based thresholding. The mm-per-pixel value was estimated using the distance between the two markers on an image, and this value was used to estimate real world coordinates for each boll and node, with the origin of the coordinate system placed at the base of the plant between the two markers. These distance coordinates were used in calculating the distances in the images, for example, the internodal distance, which was calculated as the Euclidean distance between two consecutive detected nodes on the main stalk. To remove the repeated counting of the same boll, two images taken from angles 180 degrees apart were used, and one of the images was flipped vertically by 180 degrees so that the coordinate system would align with the other image. After that the repeated counts were detected using a threshold value for intersection over union for the boll bounding boxes from the two images.

Kinect data processing

The RGB images obtained with the Kinect v2 were processed identically with the RGB images as described above. Differences included the need to crop the images for using them with the object detection network, and then the need to project those bounding boxes for the detected plant parts back to the original RGB image. This had to be done because our aim was to register the depth values from the Kinect onto the RGB image so that we could have a sparse distribution of the depth of pixels for the RGB image. This information was used to calculate the coordinates of the detected nodes and bolls in 3D space.

Results

Boll and node detection

Figure 2 shows the bolls and nodes detected on images taken from the two views that were aligned at an angle of 180 degrees to each other. The image on the left panel (*view 1)* shows a total of 13 bounding boxes for the bolls (shown in red) and 13 bounding boxes for the nodes (shown in blue) that were detected by the models, and the image on the right panel (*view 2)* shows a total of 16 bounding boxes for the bolls and 11 bounding boxes for the nodes detected by the models.

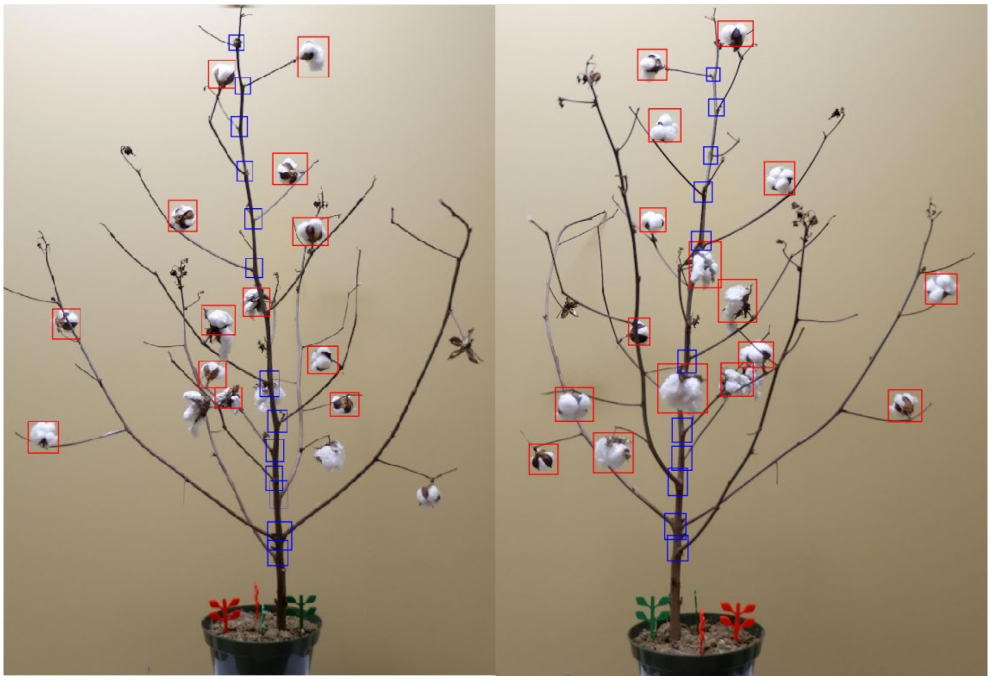

Figure 2 Images showing the locations of main stalk nodes and cotton bolls detected by the object detection model; the detected cotton bolls are shown within red bounding boxes whereas detected nodes are shown within blue bounding boxes whereas detected nodes are shown within blue bounding boxes

In figure 3, a view of the point cloud data representing the cotton plant is shown with each node and boll labelled with a number. The number shown in red circles on the main stalk is a convention for numbering the main stalk nodes that will be used in this paper. The first main stalk node with a visible branch is numbered to be "node 1", and subsequent main stalk nodes are numbered with increasing numbers. While the node with the first visible branch in the studied plant is not strictly node number one in the conventions of botanical studies (Zhao, Oosterhuis et al. 2000), the numbering adopted here is used for convenience. The bolls are given numbers according to the node number at which the branches producing the bolls originate.

For example, a boll with the number 10 beside it is produced by the fruiting branch originating at node number 10 at the main stalk.

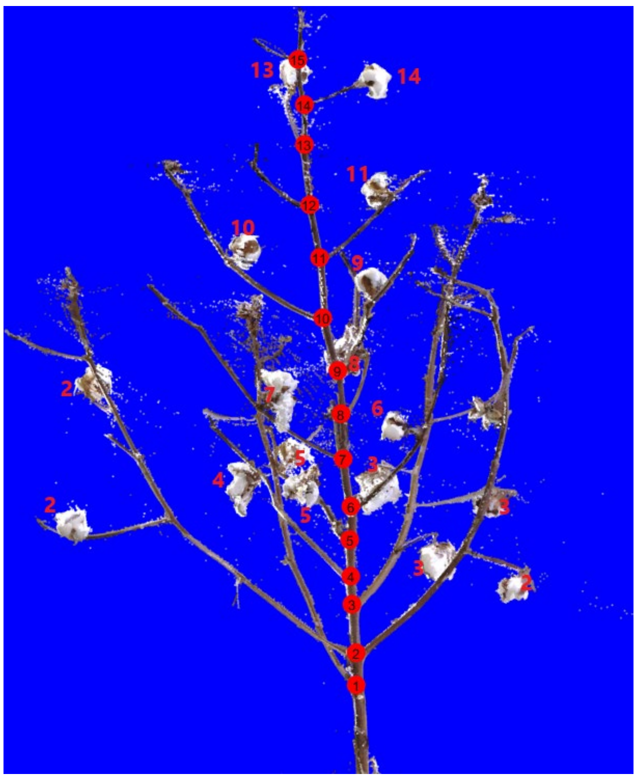

Figure 3 A view of the point cloud data of the cotton plant showing the main stalk nodes numbered from the bottom, and also showing the respective cotton bolls produced by branches originating in the main stalk node numbers shown next to the bolls; a total of 17 bolls assigned to 15 individual nodes are shown

When we compare the total number of nodes with the detected nodes (fig. 2), we see that 13 out of 15 main stalk nodes are detected and localized by the model for view 1 whereas the ratio is 11 out of 15 for view 2. Similarly, 4 bolls are missed by the detection model for view 1 and 1 boll is missed for view 2.

Internodal distance

In figure 4, we see the intermodal distances estimated from both images, and the right panel shows the ground truth distances. Figure 3 shows the plot of the estimated distances from view 1 against the ground truth distances. In creating the plot, the intermodal distances in the ground truth data were adjusted to match the detected nodes in the model derived values so that an unmatched plotting could be avoided. For example, if the model derived value sums up the intermodal distance between node 6 and node 8 and has a single value, the ground truth distances are accordingly summed so that the values are matched with one another. Figure 5 shows a plot of the internodal distances against the ground truth, where we see a coefficient of correlation value of 0.9821. Quantitative evaluation of branch angle estimation was not conducted for this study, and the visual results were shown in figure 1.

Figure 4 The internodal distances estimated using a distance per pixel value estimated using measured distance between two points in each image (left and middle); A view of the point cloud data representing the cotton plant displaying the ground truth nodes with the node numbers and internodal distances (right)

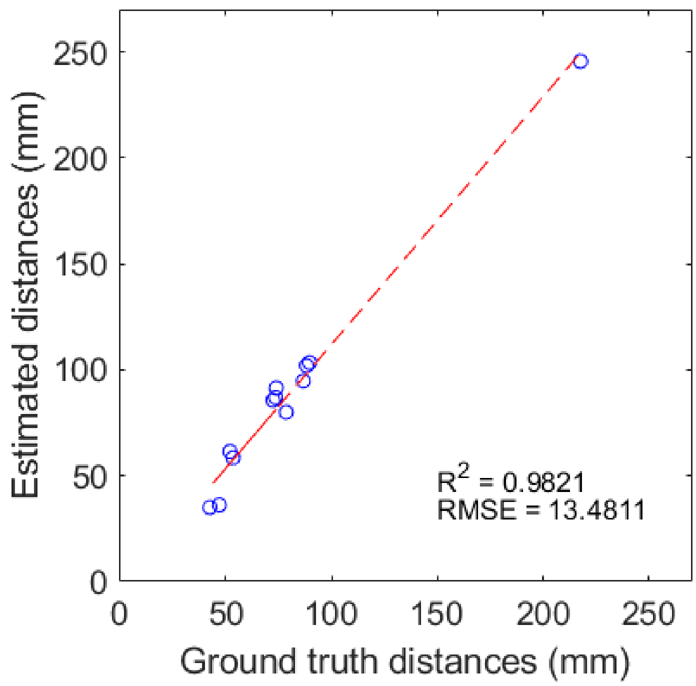

Figure 5 Internodal distances estimated from the RGB image against the ground truth

Boll assignment to main stalk nodes

In figure 6, the assignment of bolls to the different main stalk branch nodes is shown graphically. The information is presented on table 1, where we can see that although the assignment of bolls works for a majority of bolls, the total number of bolls per node estimated using this method is inaccurate for most of the nodes. This is a result of the assumption of planar geometry, where we assume that all bolls lie on one plane and simply find the shortest path to the main stalk through the branches visible on the image. This leads to errors, especially in cases where the bolls are close to the main stalk, where although the path to the main stalk seems to be short on a particular two-dimensional image, the boll may in fact be on an extended branch that grows perpendicular to the image plane. This problem can be mitigated by using depth images, where the position of bolls in three-dimensional space can be estimated, and based on similarly estimated position of the branches and the nodes, faulty assignments can be eliminated.

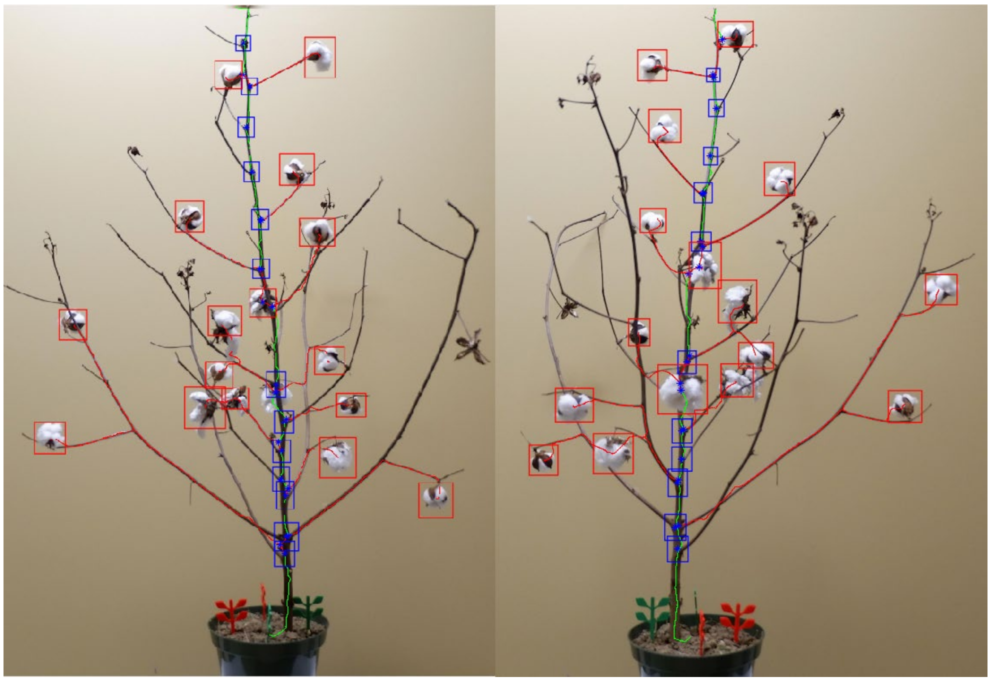

Figure 6 The assignment of bolls to main stalk nodes for view1 (left) and view2(right). The main stalk is shown with a green line and the branches connecting the bolls to the main stalk are marked in red

Another possibility for the improvement of this result is the creation of a combined algorithm to use multiple views of the plant together to have a unified assignment of the detected bolls to the main stalk nodes. The presence of main stalk nodes that are not detected by the node detection model is another issue that needs to be addressed. It is, however, easily addressed by finding the points where the path from a boll to the main stalk touches the main stalk.

Table 1 Assignment of the detected bolls to the main stalk nodes; column 2 shows the ground truth, column 3 and 4 show the result from the assignment algorithm

Node	**No. of bolls**	**Detected bolls (view1)**	**Detected bolls (view2)**
1	0	0	0
2	3	3	2
3	3	1	3
4	1	0	0
5	2	1	1
6	1	1	Not detected
7	1	3	3
8	1	Not detected	Not detected
9	1	Not detected	Not detected
10	1	3	3
11	1	1	1
12	0	0	0
13	1	0	0
14	1	2	2
15	0	0	Not detected

Boll count estimation from multiple views

In order to consolidate the bolls detected from multiple views of the same cotton plant, we need a method to recognize the same boll detected on different images, and then count the uniquely identified bolls as new bolls. This is a problem of tracking the bolls and detecting repeated counts. Using the coordinates of the bolls based on the transformation of image coordinates to real life distances described before, the bounding boxes were projected to the real world coordinates as shown in figure 7. We can see that the bounding boxes towards the top of the image, corresponding to the same bolls detected on the two views, have bounding boxes that have overlapping areas, but the bounding boxes from the same boll appear further apart towards the base of the plant, for example, at the lower left of the plot. If we assume that the bounding boxes that have an overlap correspond to the same boll, we have a total count of 19 bolls, which helps to eliminate the underestimation of boll count derived from a single image, but is still inaccurate when compared to the ground truth. However, the method of detecting repeated counts is not entirely flawed as we can see that a boll detected only on one image can be isolated from those that have repeated detection, as can be seen in case of the bolls on the lower right.

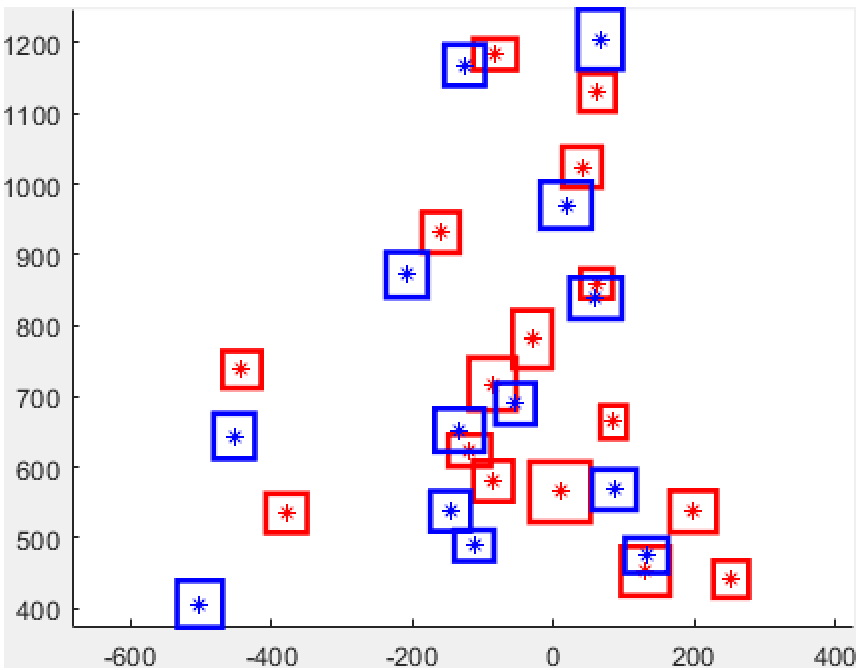

Figure 7 Boll bounding boxes from the two views of the cotton plant projected into a coordinate plane where the units are in mm. The origin of the coordinate system is assumed to be at the base of the cotton plant.

RGB and depth images from Kinect v2

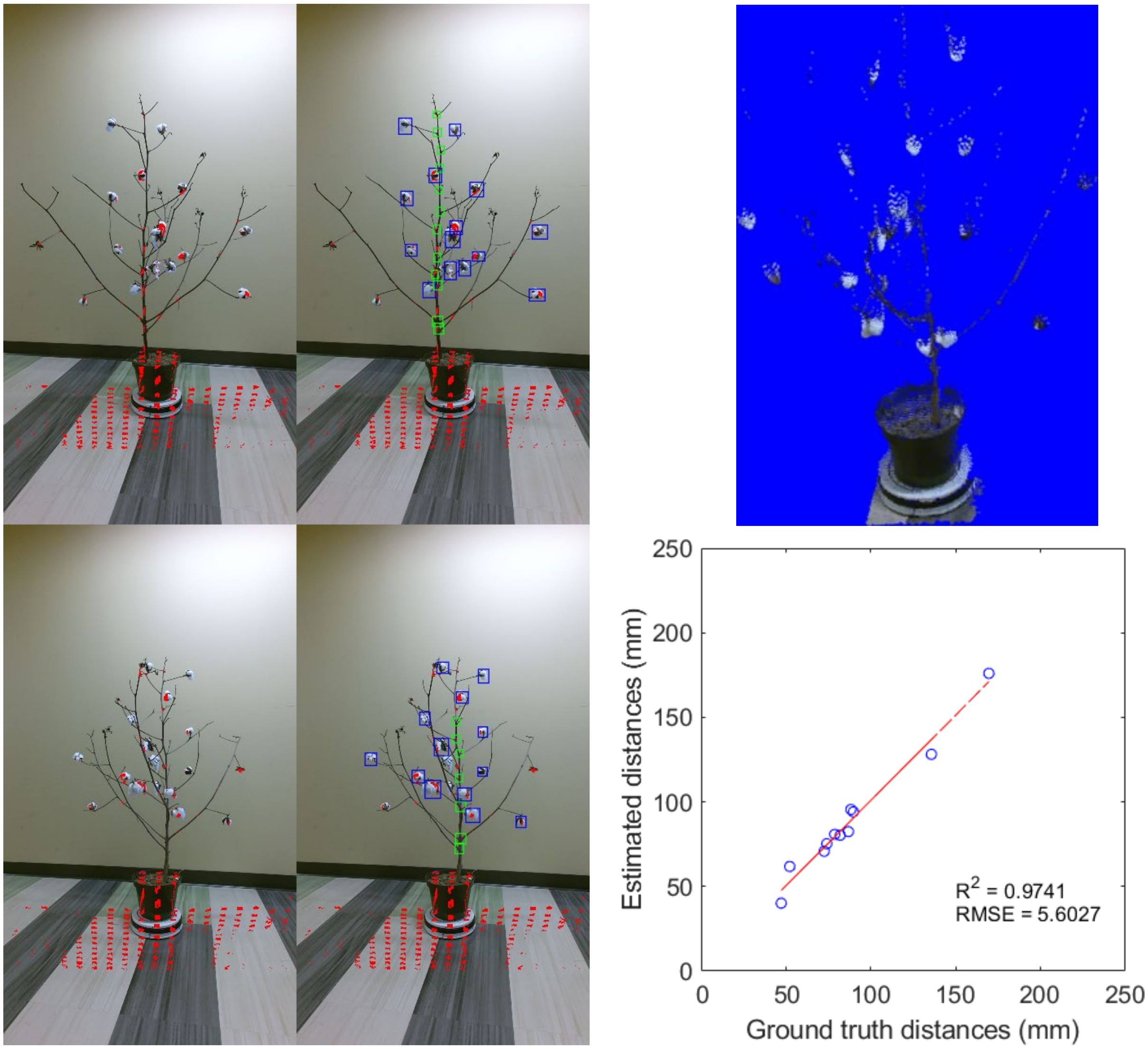

Figure 8 RGB images acquired from the Kinect v2 showing the depth values from the low-resolution depth image in red (left); nodes (green bounding boxes) and bolls (blue bounding boxes) detected on the RGB images. The two images on the upper left are from view1 and the lower two images are from view2 (180 degree apart). The average depth of a boll or node is calculated by averaging the depth values that lie within a bounding box on the projected images. The upper right image is the point cloud derived from a single depth image and the corresponding RGB image. The lower right plot shows the internodal distances estimated from the Kinect data plotted against the ground truth distances

The internodal distances calculated from the average coordinates for each node are plotted against the ground truth distances in figure 8. Here, the coefficient of correlation is comparable to the coefficient of correlation obtained using only the RGB images, but we see a significant reduction in the root mean square error, which means that the prediction accuracy of the model based on the Kinect v2 data is higher than the model based on the RGB images alone.

The process used for the detection of repeated counts in case of RGB images was also used with the Kinect data. Figure 9 shows the plot of the coordinates of the bounding boxes representing the cotton bolls detected on two images taken from two sides of the plant. The plot is on the X-Y coordinate plane, showing the superior accuracy achieved when using the Kinect data. When pairing the points according to a threshold value based on the standard deviation of the distance from the point to its nearest neighbor, we can find 16 unique bolls detected using the two views. Using a single view would provide us with 14 and 13 counts respectively for view1 and view2.

Conclusion

This paper discussed a preliminary study for the extraction of plant phenotypic parameters using deep learning methods with two-dimensional images for detection of plant parts, and the subsequent use of this information for the calculation of plant parameters. The results show that this approach can be useful in extracting a set of plant traits that are only possible with the use of three-dimensional imaging. For example, promising results were obtained for the detection of main stalk and main stalk nodes, which can be useful for the derivation of plant health and growth status. Similarly, the assignment of cotton bolls to a specific node was also attempted. The detection of plant parts and their placement in two or three-dimensional space can be useful for reconstruction of the whole plant based on this data. For example, the information on node location, branch angles, and the number of bolls per main stalk node can be used to construct a rough parametric model of a plant. The use object detection models implemented in this study can be extended to the detection of other plant parts such as branch tips and flowers. The reconstruction of the parametric model of a plant could be a useful substitute for three-dimensional scanning which is expensive both in terms of data collection and processing. We also found that the augmentation of two-dimensional RGB images using depth data can increase the accuracy of the structural traits that we attempted to derive in this study.

References

Fue, K. G., et al. (2018). Deep Learning dfbased Real-time GPU-accelerated Tracking and Counting of Cotton Bolls under Field Conditions using a Moving Camera. 2018 ASABE Annual International Meeting, American Society of Agricultural and Biological Engineers.

Furbank, R. T. and M. J. T. i. p. s. Tester (2011). "Phenomics–technologies to relieve the phenotyping bottleneck." Trends in plant science **16**(12): 635-644.

Hawkesford, M. J. and A. J. F. P. B. Lorence (2017). "Plant phenotyping: increasing throughput and precision at multiple scales." Functional Plant Biology **44**(1): v-vii.

Huang, J., et al. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. Proceedings of the IEEE conference on computer vision and pattern recognition.

Jiang, Y., et al. (2016). "High throughput phenotyping of cotton plant height using depth images under field conditions." Computers and Electronics in Agriculture **130**: 57-68.

Jin, J. and L. J. J. o. F. R. Tang (2009). "Corn plant sensing using real-time stereo vision." Journal of Field Robotics **26**(6-7): 591-608.

Kamilaris, A., et al. (2018). "Deep learning in agriculture: A survey." Computers and Electronics in Agriculture **147**: 70-90.

Li, L., et al. (2014). "A review of imaging techniques for plant phenotyping." Sensors **14**(11): 20078-20111.

Li, Y., et al. (2017). "DeepCotton: in-field cotton segmentation using deep fully convolutional network." Journal of Electronic Imaging **26**(5): 053028.

Lin, T.-Y., et al. (2014). Microsoft coco: Common objects in context. European conference on computer vision, Springer.

Pound, M. P., et al. (2017). "Deep machine learning provides state-of-the-art performance in image-based plant phenotyping." Gigascience **6**(10): gix083.

Ren, S., et al. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems.

Sun, S., et al. (2017). "In-field high-throughput phenotyping of cotton plant height using LiDAR." Remote Sensing **9**(4): 377.

Thompson, A. L., et al. (2019). "Comparing Nadir and Multi-Angle View Sensor Technologies for Measuring in-Field Plant Height of Upland Cotton." Remote Sensing **11**(6): 700.

Zaidi, S. S.-e.-A., et al. (2018). "The Rise of Cotton Genomics." Trends in plant science **23**(11): 953-955.

Zhang, Z. J. I. T. o. p. a. and m. intelligence (2000). "A flexible new technique for camera calibration." IEEE Transactions on pattern analysis and machine intelligence

Zhao, D., et al. (2000). "Dynamics of non-structural carbohydrates in developing leaves, bracts and floral buds of cotton." Environmental experimental botany **43**(3): 185-195.