Computers and Electronics in Agriculture 218 (2024) 108669

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Machine vision based plant height estimation for protected crop facilities

Namal Jayasuriya [a,*], Yi Guo [b], Wen Hu [c], Oula Ghannoum [a]

[a] Western Sydney University, Hawkesbury Institute for the Environment, Richmond, 2753, NSW, Australia
[b] Western Sydney University, Centre for Research in Mathematics and Data Science, Parramatta, 2150, NSW, Australia
[c] The University of New South Wales, School of Computer Science and Engineering, Sydney, 2052, NSW, Australia

ARTICLE INFO

Keywords:
AI for phenotyping
Stereo vision
3D point cloud
Plant height estimation
Vertically supported protected crops

ABSTRACT

The increasing demand for quality, year-round food production in limited space has led to the widespread adoption of protected cropping. Effectively monitoring and maintaining crops within these facilities requires substantial labour and expertise. Traditional manual monitoring is labour intensive and time consuming. Therefore, non-destructive image-based techniques, particularly those utilising 3D structural data, have gained attention. We developed a stereo vision-based system to estimate the height of vertically supported tall plants in protected facilities, given plant height serves as a vital measure of crop growth. Our system uses a mobile platform with a top-angle view of a stereo vision depth camera for data acquisition and machine learning in its core for data analysis. First, we collected weekly RGB and depth (RGBD) streams from plant gutters in three glasshouse compartments with different light treatments. We used part of the RGB data collected to train and validate a deep learning segmentation model to detect plant tops and bases. Detected tops and bases of an image were then mapped to the generated 3D scene using the depth image of the same frame. Thresholds and 3D clustering are used respectively to remove background and eliminate outliers in top and base detection mapped to 3D space. Finally, the height of each plant was calculated using the cluster centres of the tops and bases of the plants. Manually measured heights of ten selected plants per environment were used to validate the height estimations. Similar growing patterns were observed between imaged and manually measured plant heights, which showed strong correlations of 0.87, 0.96, and 0.79 R^2 scores, respectively, under unfiltered ambient light, Smart Glass film, and shifted light. These promising results demonstrate the feasibility of our proposed method for a vertically supported capsicum crop in a commercial-scale protected crop facility.

1. Introduction

The world population is expected to reach 9.3 billion by 2050, as predicted by United Nations Department of Economic & Social Affairs, Population Division (2022). Future food demand to feed this population is expected to increase by 62% while the risk of hunger by 30%. Increasing food production without expanding agricultural lands (Searchinger et al., 2019) and the need for year-round quality production in limited space, have increased the importance of protected cropping facilities in modern agriculture. Protected facilities offer controlled environments, allow optimal plant growth and reduce dependence on external weather conditions (Gruda and Tanny, 2014). However, advanced protected facilities have costs associated with the high energy demand of manipulating the environment conditions, the problem of pollination of crops that do not support self-pollination, and the skilled labour shortage (Chavan et al., 2022). Our overall aim is to address the skilled labour requirements. Advances in robotics, machine learning, and various imaging technologies have shown great potential in automating crop monitoring and maintenance. Here, we explored a machine vision-based automated method for estimating plant height in complex protected crop environments.

Plant height estimation is a crucial aspect of crop monitoring and phenotyping, as it provides valuable information on overall growth and production. For example, Yin et al. (2011) reported a significant correlation between plant height and corn yield at different stages of growth. Both Tilly et al. (2015) and Bendig et al. (2014) found correlations between plant height and barley crop biomass with an R^2 of 0.8. Traditionally, height measurement has been performed manually, which requires significant human effort. Image-based methods have emerged as promising alternatives, offering non-invasive and efficient approaches for height estimation and encompassing a wide range of techniques depending on crop type and infrastructure.

* Corresponding author.
E-mail addresses: N.Jayasuriya@westernsydney.edu.au (N. Jayasuriya), Y.Guo@westernsydney.edu.au (Y. Guo), wen.hu@unsw.edu.au (W. Hu), O.Ghannoum@westernsydney.edu.au (O. Ghannoum).

https://doi.org/10.1016/j.compag.2024.108669
Received 10 October 2023; Received in revised form 16 January 2024; Accepted 21 January 2024
Available online 1 February 2024

For field crops, various technologies have been employed. Wang et al. (2018) developed a high-throughput phenotyping system using a 2D LiDAR sensor for sorghum, achieving accurate plant height measurements with a high R^2 value of 0.98. Sun et al. (2017) utilised a LiDAR system mounted on a high clearance tractor for the phenotyping of cotton plants, reporting an average R^2 value of 0.98. Jiang et al. (2016) employed a Kinect-v2 camera on a high clearance tractor for estimating the height of cotton plants, achieving accuracy greater than $R^2 = 0.92$.

Madec et al. (2017) explored the use of LiDAR on an unmanned ground vehicle and a high-resolution RGB camera on an Unmanned Aerial Vehicle (UAV) for the estimation of plant height of wheat genotypes. Both techniques gave strong correlations with manual measurements with high heritability values ($H^2 > 0.88$). Sofonia et al. (2019) demonstrated that both UAV LiDAR and multispectral camera give accurate crop height estimation for sugarcane. In all these studies, LiDAR provided better results.

RGB cameras and Structure-from-Motion (SfM) techniques have been utilised in several studies. Zhu et al. (2023) proposed a methodology using a lightweight UAV and high resolution multiview images for organ-scale parameter quantification in maize, soybean, and tobacco crops, achieving high accuracy in estimating leaf phenotypic parameters. Xiang et al. (2019) also tried similar method for sorghum targeting traits including plant height achieving RMSE below 0.146 m. Reji et al. (2021) used terrestrial laser scanning (TLS) with adaptive spatial filtering, canopy height modelling, watershed segmentation, and support vector regression for plant height estimation in tomato, eggplant, and cabbage crops, achieving high prediction accuracy for various parameters.

Depth cameras have also been employed for plant height estimation. Jiang et al. (2016) developed a high-throughput phenotyping (HTP) system using Kinect-v2 camera by Microsoft for measuring the height of cotton plants, achieving accuracy of more than 92%. Morrison et al. (2021) integrated the RealSense D415 camera by Intel® RealSenseTM into the PlotCam platform to measure crop heights on wheat and soybean plots, showing significant correlations with single-point LiDAR measurements. Kim et al. (2021) found strong agreement between estimated crop heights and manual measurements for various field crops, using stereo vision and showing R^2 values ranging from 0.78 to 0.84. Another depth camera-based 3D scanning for sorghum plants was conducted by Xiang et al. (2019) for traits including plant height.

Although significant progress has been made in monitoring field crops, such as rice, corn, cotton, and maize using drones and tractors with sensor systems, monitoring vertically supported crops in protected facilities has rarely been studied. Plants tend to grow taller in protected facilities compared to open field conditions due to well-managed environmental conditions and inputs. The unique challenges posed by these crops, including taller plants, variable light conditions, limited space, the presence of support structures, and potential occlusions, require tailored approaches to provide reliable and accurate height estimation (Jayasuriya et al., 2024). We found only one study by van der Heijden et al. (2012) who used time-of-flight (ToF) and RGB cameras to 3D reconstruct tall capsicum plants in a greenhouse. They have used a trolley that moves on heating pipes on the floor with vertically stacked four camera modules, including an RGB and a ToF camera per module. A blue-coloured screen has been used behind the plant line to separate background plants and identify the top points of the canopy using pixel colour differences. Each plant base is identified using the QR code underneath the plant, and the green top point with a density of 40 pixels in the 251-pixel range centred on the QR code is considered as the plant top. They have also used flash lights to obtain sharp images with less disturbance from ambient light. Their methodology showed a correlation of 0.93 between manually measured and image-based height. However, the blue screen used for foreground background separation and canopy top identification makes it less robust for automation in a commercial setting. Also, four camera modules with a flash light make it complex and costly. Furthermore, their method of identifying the top region as the plant top does not guarantee the identification of the actual plant top.

In this study, we developed a new approach for plant height estimation consisting of several modules. First, we extract RGB and Depth images using Realsense D415, then Mask-RCNN (Region based Convolution Neural Networks) (He et al., 2017) is used to identify plant tops and bottoms, and the instance identification is fused with 3D scene obtained from depth image, and noise is filtered. Finally, height is calculated using filtered tops and bases of the plants. The neighbour frames of sliding windows are used to identify the missing plant tops and all the identified plant bases are considered to determine the base level of a gutter. See Fig. 1 for the detailed workflow.

2. Materials and methods

2.1. Plant culture and growth facilities

The experiment was conducted at the National Vegetable Protected Cropping Centre (NVPCC) at Western Sydney University's Hawkesbury campus. The NVPCC is a state-of-the-art smart glasshouse facility dedicated to research and education in protected cropping for vegetable production. The facility is described in detail by Chavan et al. (2020). Briefly, the NVPCC offers unparallelled control over microclimate conditions such as temperature, humidity, and CO2 levels. The plants grow in a soilless rockwool media with an automated system for fertigation. Seeds for Capsicum annum L. (red gina variety) were donated by Syngenta (2023), and grown in a nursery for up to six weeks before being transplanted into Rockwool slabs. The seedlings were then moved to three glasshouse rooms each covering an area of 105 m^2. The Control room incorporating advanced roof glass that diffused 70% of the light and wall glass that diffused 5% of the light. The roof of the Smart Glass (SG) room was equipped with a light blocking film (LBF) that blocks 86% of ultraviolet light, 26% of red, and 58% of far red (Chavan et al., 2020). This Smart Glass is used for the roof of the room to minimise the energy use for climate condition manipulation (Lin et al., 2022). The roof of the third room was covered with a light shifting film (LLEAF) that shifts the green spectrum to red, aiming at increased crop growth and production (Soeriyadi, 2023). Each room includes six hanging gutters (length 10.8 m, width 25 cm, AIS Greenworks, Castle Hill, Sydney, NSW, AUS) with ten Rockwool slabs per gutter (100 × 15 × 10 cm, Gordan, the Netherlands). The outer two gutters acted as buffers. The central four gutters had 4 capsicum plants per slab (0.25 m gap between two plants), which makes 40 capsicum plants per gutter. There were two main stems per plant, growing up to 2.5 m in height. Taller plants need vertical support. Glasshouse rooms also include ground pipe systems where heated water flow to increase room temperature and which also serve as a rail system to move trolleys and elevated platforms for crop management and harvesting (Fig. 2).

2.2. Camera selection

We considered Servi et al. (2021)'s comparison on three affordable depth cameras, including two stereo cameras (D415 and D455) and the D515 LiDAR scanner according to the advanced international standard that regulates the metrological characterisation of coordinate measuring systems (ISO 10360-13:2021). Although the D515 LiDAR scanner is on a low budget compared to other LiDAR scanners, it is designed to work only in indoor conditions, while the D400 series is designed to work in both indoor and outdoor conditions. They have conducted 3D reconstruction for close range and concluded better reconstruction quality from D415. Hence, we selected the Intel Real Sense D415 as a low-cost IR-based stereo camera that also comes with an integrated RGB camera and a toolkit to post-process and align RGB and depth images (Intel-RealSenseTM, 2023). Their Infra-Red (IR) based stereo

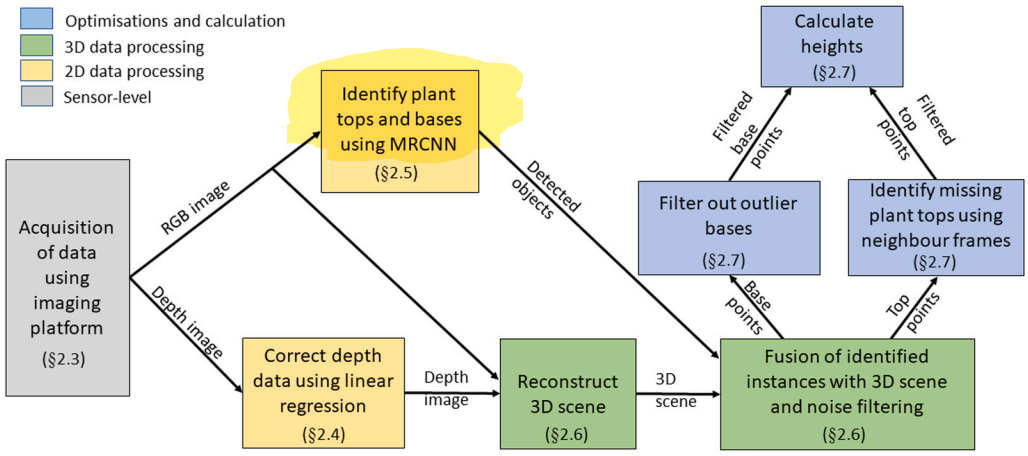


Fig. 1. Data processing pipeline of the methodology propose by this work.

Fig. 2. Interior structure of a glass house compartment.

vision has been designed to be less affected by ambient light and also to generate a better depth image under low light conditions for depth calculation, while consuming relatively low extra power (Grunnet-Jepsen et al., 2018). It suits our environment, which is affected by ambient light, even though it is an indoor environment. The target crops in this work grow up to 2.5 m, and the ideal range of Intel RealSense D415, which is 0.3 to 3 m, matches our requirement.

2.3. Data acquisition

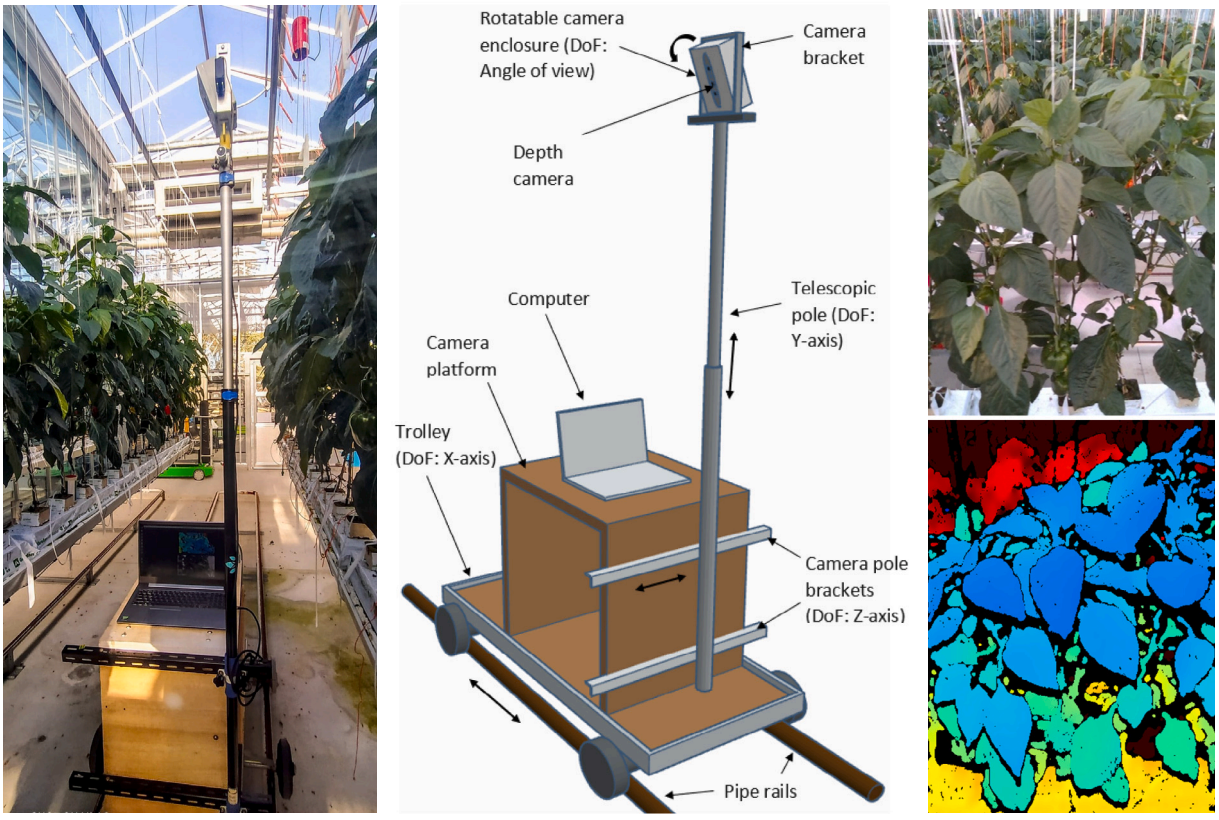
As we highlighted in the related works, there is a lack of research on automating phenotypical trait extraction for commercial scale vertically supported crops in protected facilities. We collected data primarily targeting the automation of plant height estimation for capsicum crop in the NVPCC. The third and fourth gutters (plant lines), located in the centre of the glass house rooms in NVPCC, were used for data collection for three environments with different light filters: natural diffused light (control), smart glass film (SG), and light shifting film (LLEAF). We refer to these three environments as ControlEnv, SmartEnv, and LleafEnv, respectively. Image data were collected once a week from September 2022 to January 2023 (21 weeks). Manually measured height was also collected for the right side branches of 10 plants per gutter for the same gutters on the same day. We used a mobile platform that can move on rails between two plant gutters with angled view cameras (Fig. 3). The platform was moved from one end of the gutter to the other, targeting the middle 33 of the 40 plants. We did a separate recording for each side of a gutter, targeting the two main branches of plants. In order to have a better view of the highest plant in the two chosen gutters of the room, the camera height is set at a 30-degree angle between the camera and the camera pole for each recording. The perpendicularity between the camera pole and the piperail system was ensured each time.

The Intel Realsense D415 camera was used with a laptop to collect both RGB and Depth streams as rosbag files. The camera was integrated with the latest firmware with factory calibrated status, according to the manufacturer's recommendation for new cameras. We did not use on-clip calibration in the glasshouse environment since it did not improve the quality of the depth image for our scene. We assumed that neither the SG film (which blocks UV and some IR; we used IR emitters on the camera) nor the LLEAF film (which increases some IR and can serve as a coherent IR source for the camera, as recommended by Grunnet-Jepsen et al. (2018)) would have an impact on the IR-based depth camera. Interference to the IR camera from other objects in the glass house, such as plants, rock wool, and metal objects, was assumed to be ignored. Both RGB and depth cameras were operated in the default preset mode with auto-exposure, and for depth imaging, always on IR emitters with a power of 300 were also employed. The recommended resolution for the D415 camera, 1280×720, was used for both RGB and Depth streams, and data were collected at a 15 frames per second frame rate with an average speed of 0.33 m per second.

2.4. Data prepossessing

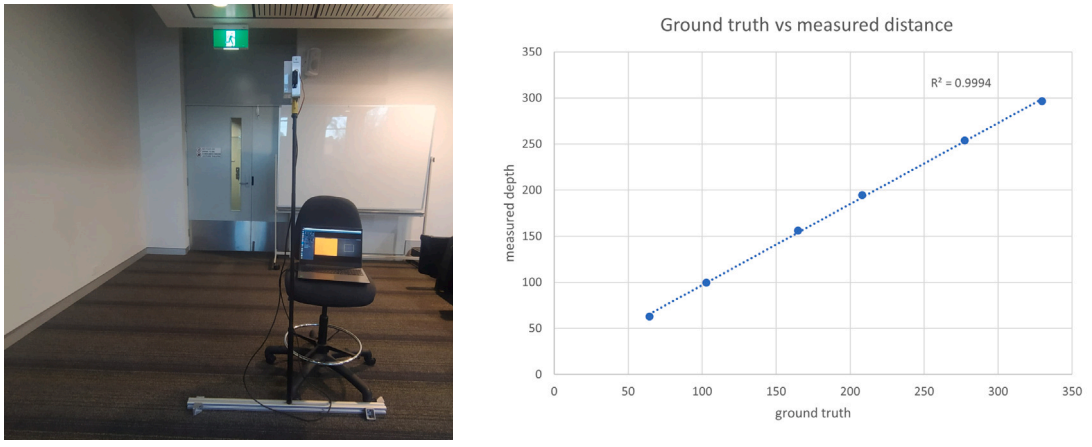
We manually selected the beginning and end of each recorded RGBD stream to choose the starting and ending plants for height estimation. Then, an offset was set to extract a frame for each plant that we needed to measure the plant height. We applied an edge-preserving smoothing filter (Gastal and Oliveira, 2011) and a decimation filter with magnitude 2 to reduce the noises of depth images and depth scene complexity. The depth image was aligned with the colour image. Despite using a relatively new camera, we later discovered that the depth estimation over distance was greater than expected due to a calibration issue, which was found to have a linear relationship with the actual distance. We used the identified linear correlation to correct the depth error.

We gathered a data set using the same camera module with the same settings in an indoor environment. The true distance from the camera plane to a wall is measured with a spin tape and the depth



(a) Imaging platform (b) Platform Schematic (c) Colour and depth images

Fig. 3. (a). Prototype imaging platform that we used for data collection, (b). 3D-schematic diagram of imaging platform and (c). sample colour image (top) and colourised depth image (bottom) obtained from this platform (distance increase from blue to red in the depth image).



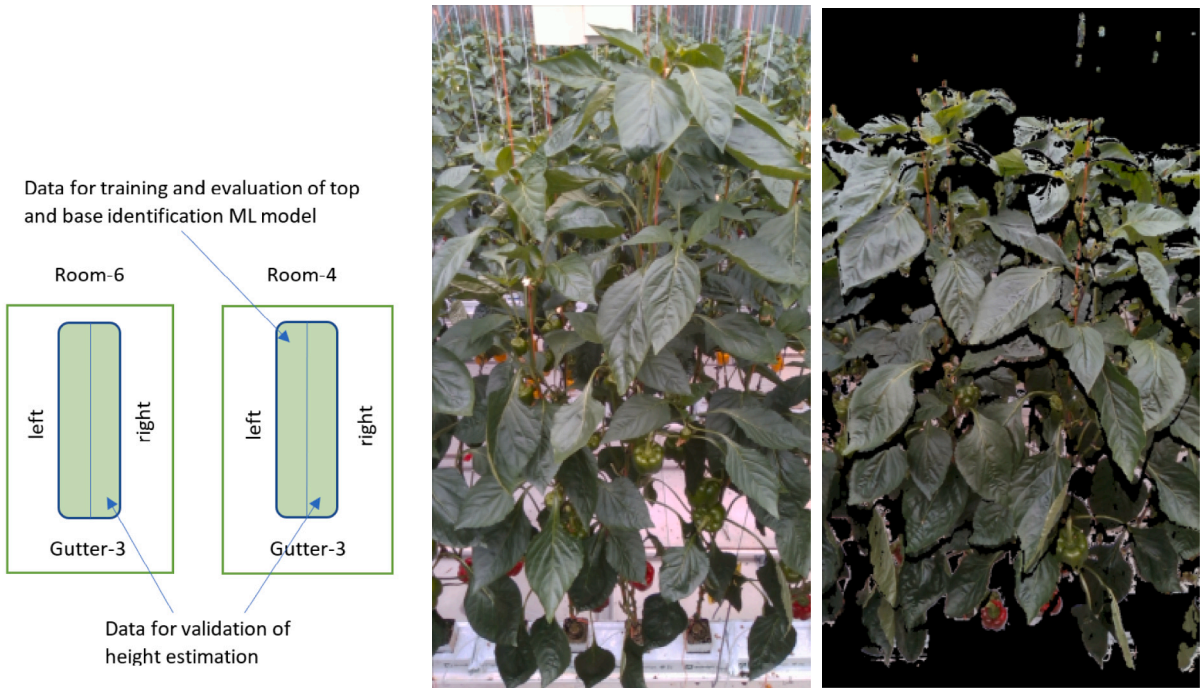
(a) Experiment setup (b) Ground truth versus calculated distance.

Fig. 4. Identification of depth error with distance. (a). Experiment setup to measure ground truth and calculated distance using depth image. (b). Correlation of ground truth height (x axis) with calculated height (y axis).

image-based distance using the Realsense D415 with the depth quality tool for a range of 0.5 to 3 m. Then the data is fitted to a linear regression model to find the model parameters. Finally, the model is used to correct the depth estimation error of the collected data for the estimation of plant height. Fig. 4 illustrates the setup of the experiment and the relationship between the calculated and true distances.

2.5. Plant tops and bases identification

Following the work of van der Heijden et al. (2012), we first tried identifying the tops of the plants using the green points of the top of an image by removing background plants using the acquired depth information and colour filtering. Single threshold depth filtering did not



(a) Data usage (b) Obstacles and vertical supports (c) Noisy green points on top of plants

Fig. 5. Data plan and problems while using top green point as plant top: (a) data usage for model training, validation and height estimation, (b) obstacles and vertically supporting wires on top of plants, and (c) noisy green points on top of plants after depth and green colour filtering.

work due to the top-angle view of our camera setup, which uses a single camera to capture the full height of plants. Hence, we used threshold filtering to remove background plants after rotating the created 3D point cloud using the measured camera angle. The problems we faced with this method are: (1) The top point was not always the top of the plant due to obstacles on top of the plants such as vertically supporting wires and sensor units. (2) Green-colour filtering to distinguish plant tops among obstacles did not work due to green-colour points from the background appearing in the foreground because of noise in the depth image. Green colour filtering is also not robust due to varying light conditions with ambient light, and it requires different thresholds according to light conditions. Fig. 5(b) shows obstacles on top of plants and vertically supporting wires, and (c) shows noisy green points on top of plants even after depth and green colour filtering. Although noise filtering was used to reduce the noise, the autonomous identification of individual plant top and base levels was still challenging. Hence, we continued with a machine learning approach to identify plant tops and bases.

Mask RCNN (MRCNN) is a state-of-the-art image segmentation architecture that has significantly advanced in the field of computer vision. MRCNN extends the Faster R-CNN architecture (Ren et al., 2015) by incorporating a mask branch for precise pixel-level segmentation of objects. This model has found applications in various domains such as object tracking, autonomous driving, and medical imaging. We recognised the tops and bases of the plants and their segmentation masks using the MRCNN provided by the Detectron2 framework (Wu et al., 2019). Detectron2 is a cutting-edge object detection and segmentation library built on PyTorch. With its modular design, flexible configuration options, and extensive model zoo, Detectron2 provides a robust and efficient framework for working with MRCNN-based applications across various platforms.

Fig. 6 presents a visual representation of the MRCNN network architecture with intermediate stages, presented using a sample image from our data set. The network uses feature pyramid networks (FPN) to extract feature maps from the input image at various scales. The initial stage involves the Stem, which captures preliminary features from RGB images of three channels and produces 64-channel feature maps. Subsequently, ResNet (residual network) based backbone network is used. In the backbone network, four residual modules are continuously applied to extract features at distinct scales. The resulting features, generated through FPN, are denoted as P2, P3, P4, P5, and P6, which respectively represent 1/4, 1/8, 1/16, 1/32, and 1/64 scales and each comprising 256 channels.

Using feature maps at these five scales, the region proposal network (RPN) functions as a Fully Convolutional Network (FCN) to identify object regions, compute objectness probabilities, and determine anchor box placements. The objectness probability signifies the likelihood of an object's presence within a region, while the anchor box specifies the region's position in the original image. The RPN ultimately delivers the top 1000 proposal boxes with the highest objectness scores to the ROI (region of interests) heads.

The ROI heads in this paper consist of a segmentation head and a bounding box head. Employing the feature maps and proposal boxes as input, the segmentation head uses ROIAlign (depicted as the purple trapezoid in Fig. 6) to generate a fixed-size 14×14 feature map. Subsequently, segmentation is performed using FCNs and a single deconvolution layer, resulting in a per-pixel segmentation map for each bounding box prediction from the box head. The box head, on the other hand, obtains a fixed-size 7×7 feature map through ROIAlign. This feature map is then converted into a one-dimensional vector, leading to the eventual output of box positions and classifications using multiple fully connected layers.

We used the pre-trained weights from the ResNet-50 model and did transfer learning with our custom data set consisting of 505 annotated plant tops and 335 annotated plant bases. To form the data set for model training and evaluation, ten random images were extracted from the data of every two weeks from the left-side branches of Gutter-3 in SmartEnv, over six months (120 images). Then, we manually annotated

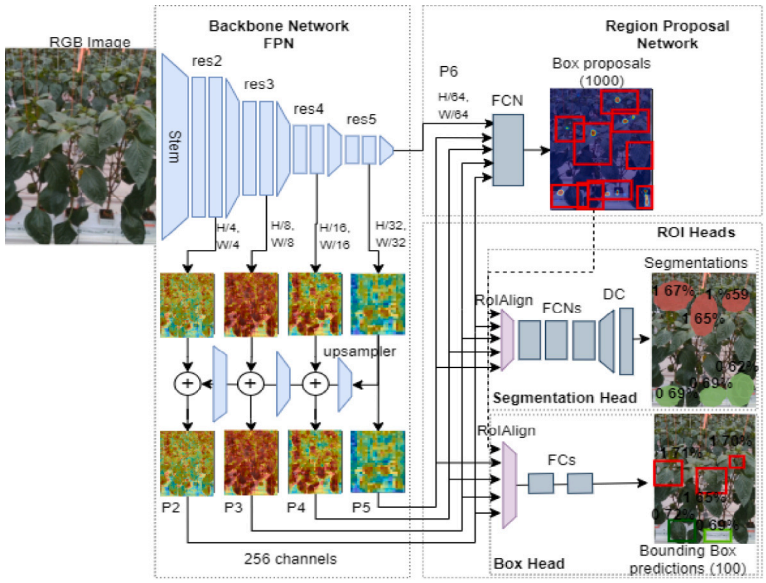


Fig. 6. Network Architecture of Mask-RCNN using a sample image to represent its stages.

Table 1
Data for training MRCNN to identify plant tops and bottoms.

Category	Training set	Evaluation set	Test set 1	Test set 2
Plant top	505	126	58	55
Plant base	335	76	36	39

Table 2
Specifications of computers used to perform data processing and measure time consumption.

Computer	CPU/cores	Speed(GHz)/ RAM(GB)	GPU/cuda cores	Memory(GB)/ bandwidth(GB/s)	Operating system
Cloud Server	Intel Xeon 4316/32	2.3/125	A100/6941	40/1555	Fedora 30
Laptop	Intel i5-10300H/8	2.5/16	GTX 1650 Ti/896	4/128	Ubuntu 22.04

Table 3
5-fold Cross Validation result of Mask-RCNN object detection and segmentation with 5800 training iterations on weights of model-Zoo ResNet-50.

Data set id	Training size	Evaluation size	Precision: Box/Seg	Recall: Box/Seg	Avg.IoA	Avg.Conf
Set 1	838	204	0.74/0.72	0.73/0.72	0.84/0.82	86
Set 2	840	202	0.69/0.69	0.77/0.77	0.85/0.84	88
Set 3	835	207	0.68/0.69	0.78/0.79	0.86/0.83	84
Set 4	831	211	0.74/0.72	0.67/0.66	0.83/0.81	86
Set 5	824	218	0.70/0.68	0.75/0.74	0.87/0.85	86
Average	834	208	0.71/0.70	0.74/0.74	0.85/0.83	86
Std.deviation	5.68	5.68	2.53/1.67	3.90/4.50	1.41/1.41	1.26

them for plant tops and bases using the CVAT annotation tool (CVAT.ai Corporation, 2022). After that, we shuffled and split the data set into five subsets for a 5-fold cross-validation of the model.

We implemented an evaluation function for MRCNN to obtain precision and recall together with the average confidence value and the average intersection of area (IoA) value per category. IoA is the intersection of prediction with the ground truth mask. IoA above 50% and confidence above 50% are considered true positive predictions. We used True Positives (TP), False Positives (FP) and False Negatives (FN) for calculating precision: $TP/(TP + FP)$ and recall: $TP/(TP + FN)$, as it does not consider true negatives with detection tasks. In other words, precision shows the true detections to total detection ratio, and recall shows the true detections to the total number of relevant instances ratio.

The validation data set, which is 20% of the data, was unseen for the model that we train each time. Another 10 images from the right-side branches of third gutter's plants were obtained in SmartEnv (test set 1) and ControlEnv (test set 2) over 5 months (random two images per fortnight) for further model testing. The use of data for training, validation, and testing of the model is shown in Fig. 5(a), while Table 1 shows the number of instances of each category for training, validation, and testing. We found better testing accuracy with learning rate 0.00025, batch size 512 per image, and L2 regularisation with 0.0001 weight decay. We also applied random lighting (0.5 probability), random contrast (scale of 0.5 to 1.5 range), and random saturation (scale of 0.5 to 1.5 range) augmentations to the training dataset. We discarded bounding boxes and segmentation with classification scores below 0.5. We selected the model at 5800 iterations as the optimal model for the detection of the tops and bases of plants for this work (model corresponds to Data set id "set 1" in Table 3). The time consumption for inference per image was measured on a general purpose laptop computer and on a cloud GPU server (specifications are shown in Table 2).

2.6. Reconstruction of 3D scene and identification of objects in 3D space for plant height

The detected tops and bases on the RGB image are needed to transfer to 3D space to calculate the height of the plants. First, we

(a) Plant top identification in sliding windows (b) Top and base detection in 2D (c) Top and base detection in 3D

Fig. 7. Use of overlapping frames and top and base detection: (a) example plant top distribution in segments of a frame and identifying the same segment in neighbouring frames, (b) plant top and base detection using the trained model on a RGB image, (c) plant top and base identification in 3D space. Red spires show the centres of the top clusters, blue spires show the centres of the base clusters, and clusters in cyan show the tops identified in the background, while clusters in orange show bases identified in the background.

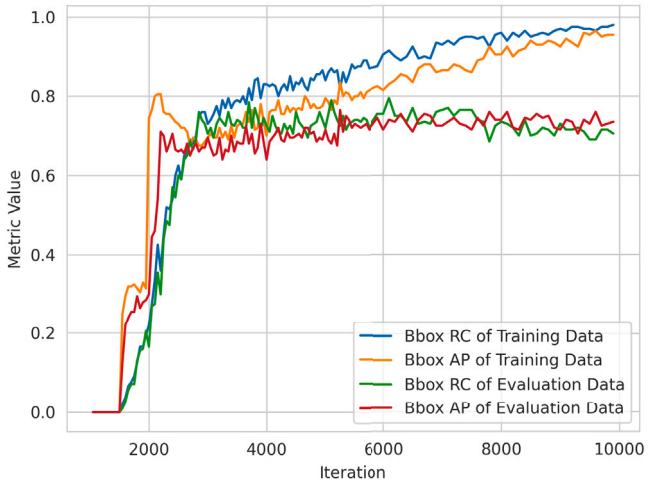


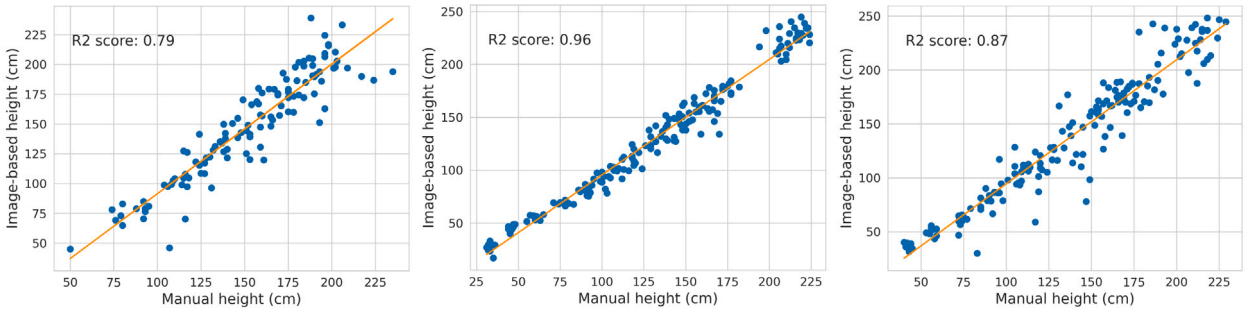
Fig. 8. Model evaluation metrics for both training data and evaluation data over number of iterations trained.

aligned the depth images with the corresponding RGB images. Then the image preprocessing and calibration error correction were carried out as mentioned in Section 2.4. The pre-processed RGB and depth images were then used to reconstruct the 3D scene using the Open3D library (Zhou et al., 2018), which is an open source library for 3D point cloud. Since the images were captured with a top-angled view, the camera angle was used to correct the orientation of the reconstructed scene, and a threshold was used to drop the background plants in the scene (plant gutters are separated with the same gap among two gutters in the green house facilities). We also used radius-based outlier filtering from the Open3D library to remove noisy voxels from the 3D scene. Then we mapped the identified segmented masks in the 2D space to the 3D space using the X and Y coordinates. Since each detection identifies a region in the 2D space, it also includes some background voxels when transferred to the 3D space. We used density-based clustering (DBSCAN) from Open3D library on identified top and base segments of 3D space and used the largest cluster as the plant's top or base. Since each plant has two stems, a threshold (perpendicular depth from camera to horizontal plane goes through the centre of a gutter along the gutter) is used to filter out plant tops identified in the background branches of the plants. To measure height, we considered the cluster centre of each identified top as the plant top point and the mean of the base cluster centres as the base level. Then we calculated the displacement on the vertical axis between each identified top point and the base level as the plant height. In Fig. 7(b) shows the tops and bases detected in the plant on an RGB image, and (c) shows the detected segments transferred to the 3D space. The top cluster centres are shown as red spires, and the bottom cluster centres are shown as blue spires, while the detected plant tops at the background branches are coloured cyan and the bases detected in background are coloured orange. Time consumption was also measured per frame on a general purpose laptop computer and on a cloud GPU server (specifications are shown in Table 2).

2.7. Sliding windows and outlier removal to improve height estimation

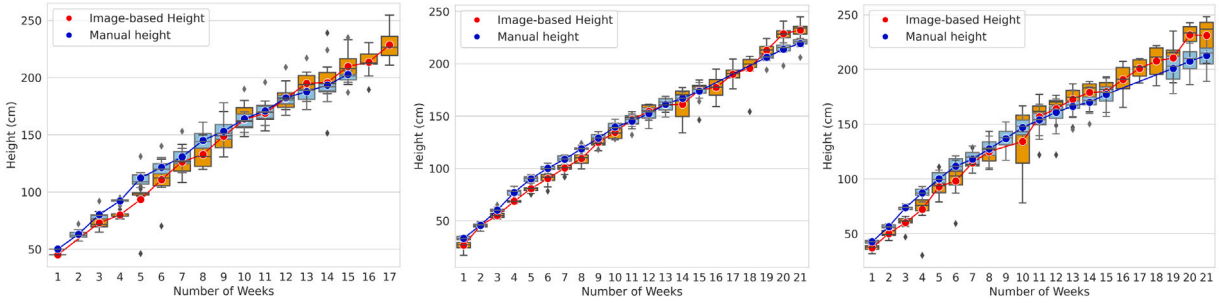
While automating the plant height estimation, we manually selected the start point and end point of the data stream as we did not have a robotic system to automatically start and end the data collection. We set an offset using the total number of frames and the number of plants captured to extract one frame per plant. Although the image data are available for the middle 33 plants, only the middle 19 plants were targeted for height estimation for each stream because the ground truth height was collected for 10 plants from the middle 19 plants. As we can see in Fig. 7(b), one frame captures three plants in the scene on average. Using this property, we segmented a frame into three regions: left, middle, and right, as visualised in Fig. 7(a). Hence, we can identify the middle plant of frame$_i$ on the left of frame$_{i+1}$ and on the right of frame$_{i-1}$. We used this sliding-window property to detect a plant when it is not detected in the middle of the frame. There is a possibility of more than one plant top in a segment of a frame; in that case, we considered the topmost plant top as the top considered for height calculation.

When the plants become tall and the crop is dense, the base of the plants may not be visible in some images. However, all plants in a gutter are at the same base level, and we calculated the mean base



(a) LleafEnv (b) SmartEnv (c) ControlEnv

Fig. 9. Correlation of height estimation using our method and manual method in LleafEnv (a), SmartEnv (b) and ControlEnv (c).



(a) LleafEnv (b) SmartEnv (c) ControlEnv

Fig. 10. Height estimation over time using our method and manual method in LleafEnv (a), SmartEnv (b) and ControlEnv (c).

point by taking into account all the bases detected in the gutter, which are within one standard deviation of the mean, for the purpose of estimating the height. We employed statistical filtering when there are more than ten bases after statistical outlier filtering. In the rare case where fewer than ten bases are detected, we used DBSCAN clustering on the coordinate values of the vertical axis and took the centre of the largest cluster as the base level. Subsequently, the heights of all the plants in all the extracted frames are calculated for a gutter. If the middle height is missing, we used neighbour frames in sliding windows to find the missing height. However, there is some possibility of identifying a nearby plant when we compare the estimated height with the manually measured height to validate the method. The time consumptions for optimisation of the base level determination of plants and neighbouring for identification of the missing tops are calculated per gutter level (for the middle 19 plants). Also, for the entire workflow, the time consumption is calculated per gutter in the two hardware specs (Table 2).

3. Results and discussion

3.1. Plant top and base detection

Fig. 8 illustrates the performance of the MRCNN image segmentation model in the number of iterations for both training and evaluation data using SmartEnv plants. Both precision and recall for unseen data remain between 70% 80% after 5250 iterations, even if the performance increases for training data. The performance for evaluation data gradually decreases after 7000 iterations. We selected the model at 5800 iterations for plant top and base detection (Dataset id: “set 1” in Table 3) and best for height estimation. We terminated the training at 5800 iterations before reaching 90% precision or recall for training data to avoid overfitting, which will heavily affect the generalisability of the trained model. As shown in Fig. 8 the evaluation metrics deviate after 5800 iterations between training and evaluation. However, there is room for improving the model simply by increasing the size of the data set so that the number of training iterations can be increased.

To validate the model, we conducted 5-fold cross validation with random shuffling and results are shown in Table 3, where the training and evaluation dataset sizes are included, with average precision, recall, and IoA values for the bounding boxes and segmentations, and the average confidence score for detection. Standard deviation of dataset sizes was 5.68 among different splits for both training and evaluation sets. Models show around 0.70 precision, 0.74 recall, and 0.85 IoA for both Bonding box and object segmentation, while it shows a confidence score of 0.86 on average across all the five splits of data. Precision values show the existence of false positives, while recall values show the existence of false negatives. The standard deviations of metrics among five models are around 2, 4, 1.41 and 1.26, respectively, for the averages of precision, recall, IoA, and confidence values. The standard deviation of precision for the bounding box is 0.86, higher than object segmentation, while it is 0.60 higher for recall of object segmentation compared to the bounding box. The average metric values around 0.70 show that the models perform relatively well, and the lower standard deviation values show that the models are not highly biased for data splits.

Since we estimated height in different environments and the trained model has converged around 0.7 of metric values for evaluation data, we created another two test sets: one from the same environment where we used data for model training (SmartEnv) and the other from a different environment (ControlEnv), even though the model has already been tested with an unseen validation dataset. We notice a 20% decrease in precision and similar recall for both plant tops and bases compared to validation data for both tests as shown in Table 4. The average IoA for true positives is slightly higher than the evaluation results for the tops of the plants, similar for the bases of the unseen gutter and slightly lower for the bases of the unseen environment. Relatively high precision drop is possible when the model is capable of identifying more plant tops in background plants and small visible parts of the bases, although we have not annotated plant tops of background

Table 4
Evaluation of MRCNN model over unseen data from same environment (Test1: different plant gutter in SmartEnv) and from a different environment (Test2: ControlEnv).

Category	Evaluation: Unseen set				Test1: Unseen gutter				Test2: Unseen environment			
	Precision Box/Seg	Recall Box/Seg	Avg.IoA Box/seg	Avg. Conf	Precision Box/Seg	Recall Box/Seg	Avg.IoA Box/seg	Avg. Conf	Precision Box/Seg	Recall Box/Seg	Avg.IoA Box/seg	Avg. Conf
Plant tops	.71/.68	.62/.60	.80/.77	.84	.51/.51	.65/.65	.94/.93	.83	.51/.50	.67/.64	.90/.89	.83
Plant bases	.76/.76	.84/.83	.88/.87	.87	.62/.57	.85/.81	.83/.84	.91	.56/.52	.84/.83	.75/.76	.88

plants and all visible parts of the plant bases in the test data set. But the recall remains relatively high or similar, because most annotated objects are detected as true positives. Both tests show similar results, except for relatively lower precision for the detection of plant bases in the unseen environment and a relatively better average IoA of true positives for test data from the same environment. Overall, the results show that the model is not heavily biased towards the data from the same environment and is considerably good for use across unseen environments.

Even though the MRCNN model detected many plant tops, we only consider the plant tops of foreground plants for height estimations, and among them, the middle plant in a frame is the target plant for height estimation. Therefore, we checked the accuracy of detecting the tops of the middle plants as shown in Table 5. The accuracy was further checked after considering the left and right plants in nearby frames when the middle plant top is not detected. Ten plants over 16 weeks (160 frames) for LleafEnv, 20 weeks (200 frames) for ControlEnv, and 21 weeks (210 frames) for SmartEnv which are the same frames that we used to estimate height in the environments, were used for this evaluation. In SmartEnv, plant top identification is improved from 80.50% to 92.38% by considering neighbour frames to find missing tops, while it increases from 70.60% to 85.00% and 75.50% to 96.00%, respectively, in LleafEnv and ControlEnv. Although randomly selected unseen data claim relatively lower top detection accuracy for both same and different environments (around 0.65 recall and 0.51 precision for tops), in terms of plant top detection for height estimation which is the top detection accuracy for middle plants remains high (average of 75.50%) and higher(average of 91.13%) with sliding window property to identify missing tops. These results also show that this method is not biased towards an environment in terms of light conditions, because of using data from a single environment for training the MRCNN model.

3.2. Height estimation in different environments

Table 6 presents the statistical measures of percentage error, absolute error, R-squared value, and root mean squared error (RMSE) of plant height estimation using our method compared to the manually measured plant height for the three environments. SmartEnv has claimed 5.43% median error (6.8 cm MAE) while, respectively, 7.74% (11.28 cm MAE) and 8.40% (10.61 cm MAE) in LleafEnv and ControlEnv. The mean errors are higher (1.4% in LleafEnv, 1.8% in SmartEnv, and 2.4% in ControlEnv) than the median errors, while the standard deviations are respectively 7.89, 6.99 and 9.54. Relatively high variance has occurred due to the presence of outliers in all three environments. SmartEnv claims best R^2 score which is 0.96. R^2 score in ControlEnv (0.87) is better than LleafEnv (0.79), even though LleafEnv is better in terms of percentage error statistics. A similar trend can be seen in the absolute median error compared to R^2.

Fig. 9 shows a graphical representation of the correlation between the manually measured height and the image-based height, along with the R^2 scores and trend lines. SmartEnv shows less variance compared to the other two environments. ControlEnv also shows a slightly high scattering around the trend line, while maintaining a better R^2 score compared to LleafEnv. A detailed representation of the data distribution at each time point and height estimation over time is shown in Fig. 10. For each time point, distributions of both manually measured height (blue) and image-based height (orange) have been plotted as box plots. The dots on the line plots are mean values for the data for each week. All three graphs have missing manually measured data points from 15th to 18th weeks. Image-based height data is missing for 2nd week and 9th week, respectively, in LleafEnv and ControlEnv. 14th week of SmartEnv and the 10th week of ControlEnv shows a slightly abnormal lower estimation due to distorted images captured with slightly higher speed of the imaging platform. Both SmartEnv and ControlEnv shows high deviation of height estimation after 19th week. Overall, SmartEnv shows less variance of data with fewer outliers, while LleafEnv shows outliers for most of the time points, and ControlEnv shows higher variance especially for the last quarter of time series.

The height of a plant is mathematically calculated using the 3D point cloud generated using the obtained depth image. RealSense D415 uses stereo IR cameras with an IR emitter to calculate the depths. Therefore, we suspect that the Smart Glass film used in SmartEnv which blocks the 26% red and 58% far red spectrum (Chavan et al., 2020) has contributed to better depth images by reducing interference for the IR-based stereo vision module of RealSense D415. Grunnet-Jepsen et al. (2018) helps us verifying this by recommending non-coherent light sources around the 850 nm spectrum bandwidth for Realsense D400 series. On the other hand, we suspect that the LLEAF film which shifts green lights towards red, also showing an increase of far red (Soeriyadi, 2023), has contributed as a non-coherent light source around 850 nm spectrum for IR-based stereo vision module to have a bit lower error statistics compared to the light diffusing roof in ControlEnv.

3.3. Comparison with similar work

Compared to the closest work of van der Heijden et al. (2012), we have claimed slightly better correlation (0.93 R^2 theirs and 0.96 R^2 ours) between the estimated and manually measured height under Smart Glass film while using a single low-cost camera, no external light, no screens for background separation, separating plant tops from opposite side branches, and detecting plant tops in a robust way. In their method, they used relatively costly 4 RGB and ToF camera modules, an external light source, an advanced trolley system to hold cameras and lights, a blue screen to separate background plants, less confidence in plant top identification, and no separating the tops from opposite side plants. Our machine vision-based height estimation method is more robust than using the green top points centred among the base for identification of the plant top. However, our methodology is less accurate (0.87 and 0.79 R^2 scores) under unfiltered scattered light and under LLEAF film compared to them.

3.4. Practical usability

In terms of the practical applicability of this method, we can compare the capability of capturing growth patterns. Our method has captured a similar pattern of average plant height over time in the three environments compared to the manually measured height as shown in Fig. 10. SmartEnv shows a continuous overestimation in height after 19th week, while ControlEnv shows from 15th week. This is possible for stereo vision based depth cameras when the distance between the camera and the object increases. However, this is not visible in LleafEnv because of unavailability of data for comparison.

Taking into account the growth per plant, a plant grew on a 10 cm average and median with a standard deviation of 4.75 cm per week, which is slightly above the error range observed as 6.8 cm median, 8.4 cm mean with 6.7 std (Table 6) in SmartEnv. The lower variance

Table 5
Accuracy of middle plant top identification and its improvement with neighbour frames to identify missing tops.

Room	Number of frames	Detected middle tops (%) without neighbouring	Detected middle tops (%) with neighbouring
LleafEnv	160	70.60	85.00
SmartEnv	210	80.50	92.38
ControlEnv	200	75.50	96.00

Table 6
Errors of plant height estimation in different environments.

Room	Error%			Absolute Error (cm)			R^2 Score	RMSE
	Median	**Mean**	**Std**	**Median**	**Mean**	**Std**		
LleafEnv	7.74	9.11	7.89	11.28	13.33	11.15	0.79	17.38
SmartEnv	5.43	7.27	6.99	6.80	8.41	6.76	0.96	10.79
ControlEnv	8.40	10.84	9.54	10.61	13.78	12.37	0.87	18.52

Table 7
Time consumption for important modules and total work flow on two hardware platforms.

	On server			On laptop		
	Median	**Mean**	**Std**	**Median**	**Mean**	**Std**
Per Frame Time Consumption considering 1140 samples (s)						
Inference	0.04	0.05	0.11	0.24	0.25	0.12
3D processing	0.97	1.09	0.59	2.00	2.19	1.34
Per Gutter (19 plants) Time Consumption considering 59 samples (s)						
Base filtering	0.00076	0.00077	0.00039	0.00039	0.00056	0.00034
Neighbouring	0.00098	0.00097	0.00010	0.00089	0.00093	0.00018
Total workflow	20.00	21.69	6.91	41.00	46.42	14.54

and better accuracy in SmartEnv are verified by the high correlation for SmartEnv in Fig. 9. In terms of usability for the entire growth cycle, our method shows results for crops up to 21 weeks (2.5 m) under SmartEnv and ControlEnv. The crop has reached up to 2.5 m height by 24 weeks under all three environments and the manually measured data have been recorded. Our method is validated up to 90% of capsicum crop growth cycle in terms of height and shows better accuracy for 90% (2.25 m) of crop growth under Smart film and up to 80% (2 m) of crop growth for LleafEnv and ControlEnv. Therefore, our method can be considered to determine the growth rate under SmartEnv, and further improvements are needed to increase per plant height estimation accuracy, especially under LleafEnv and ControlEnv.

The time consumption of data processing is another aspect that needs to be considered for the practical usage of the system even though plant growth monitoring is not a time-critical task. This methodology consumed on average 21.69 s with 6.91 standard deviation for per gutter (19 plants) height estimations (full process after image frame extraction) on a high performance cloud server with a GPU cluster, while it was doubled on a general purpose laptop computer with a small GPU (see the specifications of computation platforms in Table 2 and time consumptions in Table 7). Since plant growth is measured weekly, this is fast enough for near-real-time data processing. Furthermore, per module time consumption is checked for important modules of the process. Per gutter processing time for base filtering and neighbouring for missing top identification in 3D space with height estimation were less than a millisecond on both computation platforms. The per frame inference time was less than 0.1 s on the high performance cloud server while it was around 0.25 s on the laptop computer. 3D processing was claimed to take around one second per frame on the server and two seconds on the laptop, marking the bottleneck of time consumption for the process. 3D processing included building a 3D point cloud and clustering to determine the main 3D point cluster from the identified objects in the 3D space. 3D clustering consumes a relatively longer time and increases for a frame with the number of instances identified and their sizes. Optimisations or alternatives are needed to process 3D point cloud for real-time monitoring of plant height while the imaging platform is driving on the rails.

4. Conclusion and future works

Reducing skilled labour cost which is one of major challenges in commercial scale protected crop facilities is the ultimate target of this work. Our literature study reveals the use of different camera technologies, reconstruction algorithms in both open fields and protected crops targeting relatively tall plants. We uncover the clear gap in addressing the challenges of image-based crop monitoring in protected facilities, especially for vertically supported taller crops. Similar work conducted for a similar infrastructure and similar crop (van der Heijden et al., 2012) was discussed in detail in Section 1, and we propose our methodology to monitor plant height in a robust way at a low cost. Designing a simple and affordable imaging platform to measure vertically supported taller plants, correction of erroneous depth data due to camera calibration error, identifying plant tops and bases on both 2D and 3D spaces, identifying unidentified plant tops in the middle of a frame using neighbour frames, distinguishing foreground plant tops from background plat tops, eliminating outlier plant bases, and estimating plant heights are the challenges that we addressed in this work. Table 8 summarises these challenges, the way we solved them, and other possible options. It is evident that the system we proposed is a cost-effective and viable solution for monitoring the growth of protected crops with Smart Glass film (R^2 = 0.96), LLEAF film (R^2 = 0.79), and diffuse light roofs (R^2 = 0.87). This study has confirmed the accuracy of the system to measure the growth of plant lines for 80% of the growth cycle, with the Smart Glass film showing the best results for 90% of the growth cycle and the accuracy per plant level above the noise level with a 0.96 correlation of the R-squared. If a more advanced depth camera is used, which is not affected by light conditions, the accuracy and usability of the system could be further improved for any lighting situation.

Identification of plant top and base can be improved by increasing dataset size for MRCNN model, improving MRCNN architecture, and also integrating depth images. Key point detection to plant identification also helps in plant top and bottom identification. In future studies, we plan to use object or key point tracking to track plants for better plants registration. Interference to IR-based camera due to

Table 8
A summary of challenges that we address in this work.

Challenge	How we address	Other possible options
Imaging tall vertically supported plants in protected facilities.	Using a depth and RGB camera module mounted on a telescopic pole on a trolley runs on rails, with a top-angle view.	- Gantry system to capture top angled view: need one gantry per two rows due to obstacles that give vertical support for plants. - Fixed cameras mounted on ceiling: need many to cover whole room due to less space between ceiling and plant tops when plants are fully grown. - Both are less extendable to capture organs at the middle and lower level of plants.
Correct data set with erroneous depth due to calibration issue	Use a regression model fitted with ground truth and measured depth with same erroneous camera	
Identification of plant top and bottom	MRCNN to instance segment top and bottom of plants using RGB images	- Using top green point in 3D space: cannot distinguish top, and do robust with noisy depth images and varying light conditions. - Robotic system/images to scan from bottom to top: still need to distinguish plant tops and other objects(background plants, sensors, supporting wires, cables) at plant top level.
Identifying plant tops and bases in 3D space	detect and segment tops and bases with RGB images. Then transfer the segmentation to the 3D space.	- Train a model with 3D annotations to identify segments directly in 3D space. - Train a model to detect key points on the plant stem to identify the base and top; this is very challenging because of occlusion and the view from the top angle.
Noises due to transferring 2D segmentation to 3D space	Density-based clustering of 3D points to remove outlier points and small clusters mapped to the background	
Eliminate plant tops of background plants and opposite side branches	Threshold-based filtering	
Occluded plant bases	Use nearby visible plant bases to get base level	- Extending imaging platform in this work with a side view camera
Eliminating noisy plant bases	- Statistical outlier removal - Cluster base outlier removal for a smaller number of bases	physical height of camera can be used when all plant bases are on same level.
Identifying plants	Identify plant tops instead of the full plant for height estimation. Use the plant top identified in the middle of a frame as the targeting plant, use overlap with sliding windows to identify missing plant tops	- Deep learning/Image processing to track plant tops - 3D reconstruction of the whole gutter and 3d segment organs - Using a robotic system with this work to image plants and register with manual locations for better registration
Identifying missing plant tops at segmentation	Use segmentation in overlapped frames to identify the same top	- Deep learning/Image processing to track plants or plant tops

different objects in a glass house environment can be studied to improve the methodology. Experimenting with a different camera or using sophisticated smoothing techniques to reduce depth image noise would increase accuracy without IR light filtering. The above mentioned future works or more work will be needed to improve the accuracy of per plant height estimation under Smart-film and especially for the LLeaF film and without any filter. Further, we hope to extend our methodology to measure other crops and other crop traits. Optimising 3D point cloud processing and the full process to work as a real-time system is another aspect of future research.

CRediT authorship contribution statement

Namal Jayasuriya: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Yi Guo:** Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing, Validation. **Wen Hu:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing, Validation. **Oula Ghannoum:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Wen Hu and Oula Ghannoum reports financial support was provided by Future Food Systems Cooperative Research Centre and WBS Technologies.

Data availability

Data will be made available on request.

Acknowledgements

We want to acknowledge the following people for their valuable contributions to our research: Distinguished Professor David Tissue, who leads the research centre at NVPCC; Dr. Sachin Chavan, who collaborated with us on the same crop leading the light treatment research for crops; Dr. Wei Liang, for managing the crop at the facility;

Goran Lopaticki, for managing the greenhouse; Norbert Klause and Mohammad Babla, for collecting crop height data manually; and other workers for maintaining the crops at NVPCC.

Funding

Research support for this study was provided by a project entitled "IoT for Indoor Farming" funded by the Future Food Systems Cooperative Research Centre (FFSCRC), Australia and WBS Technologies awarded to Prof Wen Hu and Prof Oula Ghannoum. NJ was supported by a scholarship funded by the FFSCRC and Western Sydney University.

References

Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., Bareth, G., 2014. Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging. Remote Sens. 6 (11), 10395–10412.

Chavan, S.G., Chen, Z.H., Ghannoum, O., Cazzonelli, C.I., Tissue, D.T., 2022. Current technologies and target crops: A review on Australian protected cropping. Crops 2 (2), 172–185.

Chavan, S.G., Maier, C., Alagoz, Y., Filipe, J.C., Warren, C.R., Lin, H., Jia, B., Loik, M.E., Cazzonelli, C.I., Chen, Z.H., et al., 2020. Light-limited photosynthesis under energy-saving film decreases eggplant yield. Food Energy Secur. 9 (4), e245.

CVAT.ai Corporation, 2022. Computer Vision Annotation Tool (CVAT). URL https://github.com/opencv/cvat.

Gastal, E.S., Oliveira, M.M., 2011. Domain transform for edge-aware image and video processing. In: ACM SIGGRAPH 2011 Papers. pp. 1–12.

Gruda, N., Tanny, J., 2014. Protected crops. Horticult.: Plants People Places, Vol. 1: Prod. Horticult. 327–405.

Grunnet-Jepsen, A., Sweetser, J.N., Winer, P., Takagi, A., Woodfill, J., 2018. Projectors for Intel Realsense™ Depth Cameras d4xx. Intel Support, Interl Corporation, Santa Clara, CA, USA.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2961–2969.

Intel-RealSenseTM, 2023. Intel RealSense D400 series datasheet. URL https://www.intelrealsense.com/wp-content/uploads/2023/03/Intel-RealSense-D400-Series-Datasheet-March-2023.pdf. Revision 015, Document Number: 337029-013.

Jayasuriya, N., Guo, Y., Hu, W., Ghannoum, O., 2024. Image based crop monitoring technologies in protected horticulture: a review. arXiv:2401.13928.

Jiang, Y., Li, C., Paterson, A.H., 2016. High throughput phenotyping of cotton plant height using depth images under field conditions. Comput. Electron. Agric. 130, 57–68. http://dx.doi.org/10.1016/j.compag.2016.09.017, URL https://linkinghub.elsevier.com/retrieve/pii/S0168169916304586.

Kim, W.S., Lee, D.H., Kim, Y.J., Kim, T., Lee, W.S., Choi, C.H., 2021. Stereo-vision-based crop height estimation for agricultural robots. Comput. Electron. Agric. 181, 105937. http://dx.doi.org/10.1016/j.compag.2020.105937, URL https://linkinghub.elsevier.com/retrieve/pii/S0168169920331422.

Lin, T., Goldsworthy, M., Chavan, S., Liang, W., Maier, C., Ghannoum, O., Cazzonelli, C.I., Tissue, D.T., Lan, Y.-C., Sethuvenkatraman, S., et al., 2022. A novel cover material improves cooling energy and fertigation efficiency for glasshouse eggplant production. Energy 251, 123871.

Madec, S., Baret, F., de Solan, B.t., Thomas, S., Dutartre, D., Jezequel, S., Hemmerlé, M., Colombeau, G., Comar, A., 2017. High-throughput phenotyping of plant height: Comparing unmanned aerial vehicles and ground LiDAR estimates. Front. Plant Sci. 8, URL https://www.frontiersin.org/articles/10.3389/fpls.2017.02002.

Morrison, M.J., Gahagan, A.C., Lefebvre, M.B., 2021. Measuring canopy height in soybean and wheat using a low-cost depth camera. Plant Phenome J. 4 (1), e20019.

Reji, J., Nidamanuri, R.R., Ramiya, A.M., Astor, T., Wachendorf, M., Buerkert, A., 2021. Multi-temporal estimation of vegetable crop biophysical parameters with varied nitrogen fertilization using terrestrial laser scanning. Comput. Electron. Agric. 184, 106051. http://dx.doi.org/10.1016/j.compag.2021.106051, URL https://linkinghub.elsevier.com/retrieve/pii/S0168169921000697.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, vol. 28.

Searchinger, T., Waite, R., Hanson, C., Ranganathan, J., Dumas, P., Matthews, E., Klirs, C., 2019. Creating a sustainable food future: A menu of solutions to feed nearly 10 billion people by 2050. Final report.

Servi, M., Mussi, E., Profili, A., Furferi, R., Volpe, Y., Governi, L., Buonamici, F., 2021. Metrological characterization and comparison of d415, d455, l515 realsense devices in the close range. Sensors 21 (22), 7770.

Soeriyadi, A., 2023. Luminescent-light emitting agricultural films. Online. URL https://lleaf.com/.

Sofonia, J., Shendryk, Y., Phinn, S., Roelfsema, C., Kendoul, F., Skocaj, D., 2019. Monitoring sugarcane growth response to varying nitrogen application rates: A comparison of UAV SLAM LiDAR and photogrammetry. Int. J. Appl. Earth Obs. Geoinf. 82, 101878.

Sun, S., Li, C., Paterson, A.H., 2017. In-field high-throughput phenotyping of cotton plant height using lidar. Remote Sens. 9 (4), 377. http://dx.doi.org/10.3390/rs9040377, URL https://www.mdpi.com/2072-4292/9/4/377. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

Syngenta, 2023. Syngenta Australia. Online. URL https://www.syngenta.com.au/.

Tilly, N., Aasen, H., Bareth, G., 2015. Fusion of plant height and vegetation indices for the estimation of barley biomass. Remote Sens. 7 (9), 11449–11480.

United Nations Department of Economic & Social Affairs, Population Division, 2022. World population prospects 2022: Summary of results.

van der Heijden, G., Song, Y., Horgan, G., Polder, G., Dieleman, A., Bink, M., Palloix, A., van Eeuwijk, F., Glasbey, C., 2012. SPICY: towards automated phenotyping of large pepper plants in the greenhouse. Funct. Plant Biol. 39 (11), 870–877.

Wang, X., Singh, D., Marla, S., Morris, G., Poland, J., 2018. Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. Plant Methods 14 (1), 1–16.

Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R., 2019. Detectron2. https://github.com/facebookresearch/detectron2.

Xiang, L., Bao, Y., Tang, L., Ortiz, D., Salas-Fernandez, M.G., 2019. Automated morphological traits extraction for sorghum plants via 3D point cloud data analysis. Comput. Electron. Agric. 162, 951–961. http://dx.doi.org/10.1016/j.compag.2019.05.043, URL https://www.sciencedirect.com/science/article/pii/S0168169919301462.

Yin, X., McClure, M.A., Jaja, N., Tyler, D.D., Hayes, R.M., 2011. In-season prediction of corn yield using plant height under major production systems. Agron. J. 103 (3), 923–929.

Zhou, Q.Y., Park, J., Koltun, V., 2018. Open3D: A modern library for 3D data processing. arXiv preprint arXiv:1801.09847.

Zhu, B., Zhang, Y., Sun, Y., Shi, Y., Ma, Y., Guo, Y., 2023. Quantitative estimation of organ-scale phenotypic parameters of field crops through 3D modeling using extremely low altitude UAV images. Comput. Electron. Agric. 210, 107910. http://dx.doi.org/10.1016/j.compag.2023.107910, URL https://linkinghub.elsevier.com/retrieve/pii/S0168169923002983.