The added value of 3D point clouds for digital plant phenotyping – A case study on internode length measurements in cucumber

Article *in* Biosystems Engineering · August 2023

DOI: 10.1016/j.biosystemseng.2023.08.010

CITATIONS
3

READS
208

3 authors:

Frans Boogaard
Wageningen University & Research
5 PUBLICATIONS **82** CITATIONS
SEE PROFILE

E.J. Van Henten
Wageningen University & Research
325 PUBLICATIONS **7,078** CITATIONS
SEE PROFILE

Gert Kootstra
Wageningen University & Research
99 PUBLICATIONS **2,060** CITATIONS
SEE PROFILE

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

The added value of 3D point clouds for digital plant phenotyping – A case study on internode length measurements in cucumber

Frans P. Boogaard [a,b,*], *Eldert J. van Henten* [a], *Gert Kootstra* [a]

[a] *Wageningen University & Research, Agricultural Biosystems Engineering Group, P.O. Box 16, 6700 AA, Wageningen, the Netherlands*

[b] *Rijk Zwaan Breeding, Eerste Kruisweg 9, 4793 RS, Fijnaart, the Netherlands*

ARTICLE INFO

Article history:
Received 28 December 2022
Received in revised form
7 July 2023
Accepted 16 August 2023

Keywords:
Digital plant phenotyping
Plant architecture
Internode length
3D point cloud
Plant-part segmentation
Deep learning

Computer-vision based methods contribute to the availability of high-quality phenotypic datasets. Most computer-vision based methods for plant phenotyping are based on analysis of 2D images. However, previous research showed that for traits related to plant architecture, like internode length, a main limitation of 2D methods was that plants with a curved growing pattern could not be accurately measured. In this work, it was hypothesised that methods based on 3D data can overcome this limitation, while increasing the overall accuracy of the internode length measurements. To test the hypothesis, a method was proposed to estimate internode lengths from 3D point clouds of cucumber plants. First, a deep neural network based on PointNet++ was trained to segment the point clouds into plant parts. The points that were predicted as 'node' were then selected and a clustering algorithm was used to group points belonging to the same node. The Euclidean distance between the detected nodes was used as an estimate of the internode length. The results were compared to the results of a previously published method based on 2D images. The results of the 3D method were significantly more accurate than the results of the 2D method. Moreover, in contrast to the 2D method, the internode length estimates of the 3D method were equally accurate for curved plants as well as for straight plants. The results clearly demonstrated that computer-vision based methods to measure plant architecture in general, and more specifically to measure internode length, greatly benefit from the availability of 3D data.

1. Introduction

Plants grown in a commercial setting are subject to requirements set by plant growers, retailers, and consumers. For example, plants should have a high resistance against diseases and pests, produce tasty fruits and have a high yield at the right time. These characteristics of the plant are determined by the genetics of the plant, and by the environment in which the plant is growing, including crop

* *Corresponding author*. Wageningen University & Research, Agricultural Biosystems Engineering Group, P.O. Box 16, 6700 AA, Wageningen, the Netherlands.
E-mail address: f.boogaard@rijkzwaan.nl (F.P. Boogaard).

https://doi.org/10.1016/j.biosystemseng.2023.08.010

Nomenclature

Point features

r, g, b	Spectral features: red, green, and blue
x, y, z	Geometric features (x, y, z location of point)

Evaluation

s_i	Ground-truth internode length between node i and node $i+1$ [mm]
$\hat{s}_i$	Estimated internode length between node i and node $i+1$ [mm]
e_i	Error between estimated internode length and ground-truth internode length [mm]
RMSE	Root Mean Square Error [mm]
IoU_c	Intersection-over-Union for class c
TP, FP, FN	True positive, False positive, False negative
P, R	Precision, Recall
F1	F1-score, the harmonic mean of precision and recall

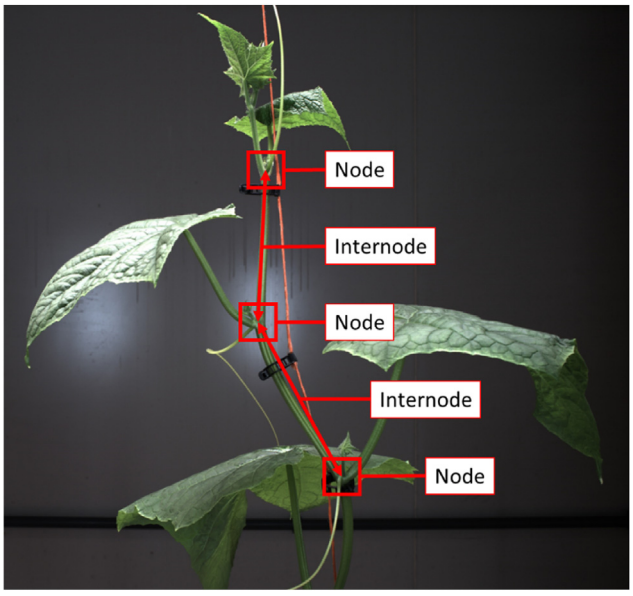

Fig. 1 – Example of three nodes of a cucumber plant with the corresponding two internodes.

management activities like de-leafing or harvesting. Plant breeding companies are developing new varieties, that are able to meet the set requirements.

For an efficient plant breeding process, breeders need to unravel the interaction between genotype, phenotype, and environment. Studying the genotype–phenotype–environment interaction requires a detailed registration of all three of these components. In this work, the focus is on measuring the phenotype of the plant. Traditionally, measuring the phenotype of the plant, called phenotyping, is done based on human observation. This is a time-consuming process (Gehan & Kellogg, 2017) and the obtained data is often subjective. In contrast, automated digital phenotyping methods open up opportunities to obtain large-scale phenotypic data in an objective way (Costa et al., 2019; Tripodi et al., 2022). Simple traits can be measured from 2-dimensional (2D) images, but measuring more complex traits, like an accurate estimation of internode length, leaf size or branching angle, requires using 3-dimensional (3D) data (Boogaard et al., 2020; Minervini et al., 2015).

The 3-dimensional organisation of plant parts, known as plant architecture (Reinhardt & Kuhlemeier, 2002), is defined by a set of variables like plant height, leaf or branching angle, leaf size, and internode length. Plant architecture has an effect on the amount of intercepted light and consequently also on photosynthesis and plant productivity. Furthermore, changes in plant architecture can indicate plant stress. Changes in internode length for example, have been related to stress caused by drought and salinity (Litvin et al., 2016; Najla et al., 2009; Sibomana et al., 2013). Internode length is the distance between two consecutive nodes along the stem of the plant. An example of some of the nodes and internodes of a cucumber plant is shown in Fig. 1. Early observation and quantification of the stress response helps to breed for more tolerant varieties. Therefore, measurement of internodes length was taken as a use case.

In earlier work, an automated method to measure internode length in cucumber plants was presented (Boogaard et al., 2020). This method was based on 2-dimensional images. These images were taken from multiple viewpoints around the plant. In each image, the nodes were detected and the detections from the different viewpoints were then combined. Based on the combined detections, the Euclidean distance between the nodes was taken as an estimate for the internode length. For the conversion from pixels to mm, a fixed plant-camera distance was assumed. However, it turned out that for 25% of the plants, the curves of the stem led to a significant violation of this assumption. The difference between straight and curved plants is further analysed in section 2.1. The reported mean absolute error of the estimated internode lengths for plants that had a straight growing pattern was 7.7 mm, while the mean absolute error for the curved plants was 23.0 mm. This suggested that accurate internode-length estimation could only be achieved when 3D data was used, as in that case, the fixed plant-camera distance would no longer be required.

Therefore, in the current work, the hypothesis was that *"using 3D point clouds instead of 2D images allows to estimate internode lengths of curved plants with the same accuracy as obtained for straight plants"*, and furthermore, that *"the error of the internode lengths estimated from the 3D point clouds is smaller than the error of the internode lengths estimated from the 2D images"*.

Contributions of the paper: To test the two hypotheses, the 3D point cloud segmentation method of Boogaard et al. (2021, p. 2022)was extended. The previously published method was developed to segment point clouds of plants into plant parts. In the current work, the previously published dataset and method were combined, to obtain segmented point clouds for all plants in the dataset. Based on the segmented point clouds, a node instance detection algorithm is proposed, to identify how the segmented node points could be grouped, resulting in a set of node objects.

Based on the detected node objects, the internode length was then determined as the distance between two consecutive nodes along the stem of the plant. To calculate this, for each node it was determined whether the next node of the plant was also detected. If that was the case, the corresponding internode length was estimated as the Euclidean distance between the two nodes in 3D space, similar to how the internode length was estimated in the 2D method in 2D space.

The first hypothesis was tested by comparing the 3D internode lengths for straight and curved plants. For the second hypothesis, the internode lengths of the same plants were also estimated based on the 2D method of Boogaard et al. (2020). The 2D internode length estimates were then compared to the 3D internode length estimates.

In the remainder of this paper, first, the materials and methods are presented in section 2, starting with an overview of the dataset in section 2.1. The 2D method is summarised in section 2.2 and the 3D method is presented in section 2.3. The evaluation methods are then introduced in section 2.4. The results of the point-cloud segmentation, clustering algorithm, node detection, and internode length estimation are presented in section 3. Finally, the results are discussed in section 4 and in the conclusion in section 5, a reflection on the hypotheses that were presented above is presented.

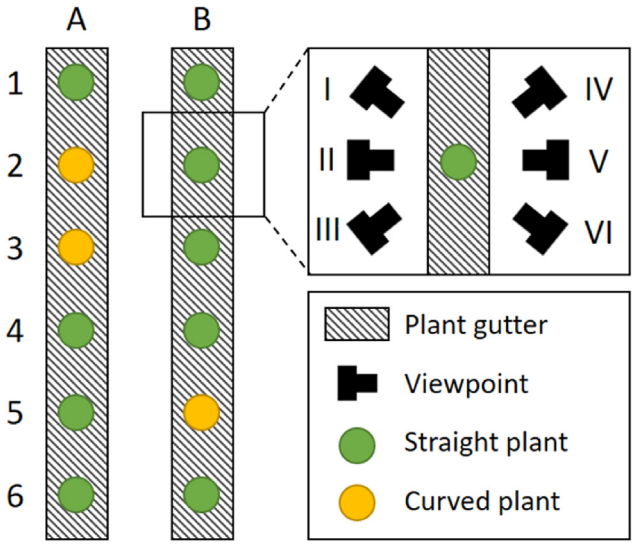

Fig. 2 – Schematic top view of the experimental set-up. The plants were grown on two plant gutters (A and B), containing six plants (1–6) each. The curved plants A2, A3 and B5 were referred to as outlier plants 4, 5 and 8 in (Boogaard et al., 2020). The six viewpoints (I-VI) are shown in the zoomed in part at the top right corner of the Figure.

2. Materials & methods

In this section, the materials and methods are presented, starting with the data acquisition and experimental set-up in section 2.1. The 2D method to estimate internode lengths is summarised in section 2.2. The approach to estimate the internode lengths from the 3D point clouds is explained in section 2.3. Finally, in section 2.4, the evaluation methods are presented.

2.1. *Data acquisition and experimental set-up*

Twelve cucumber plants (Proloog RZ F1, Rijk Zwaan, De Lier, The Netherlands) were placed on plant gutters with a plant distance of 1 m to prevent occlusion between plants. A schematic overview of the plants is provided in Fig. 2.

The data acquisition was done using a Wiwam plant analyser[1] equipped with a 2D camera used for the 2D internode length estimation and a Phenospex F500 dual scan system[2] for the 3D internode length estimation. The 2D camera that was used is an IMPERX B4820 16 MP CCD camera with an image resolution of 4904 × 3280 pixels (IMPERX, 2018). The 3D data was based on laser-line triangulation and was collected using a Phenospex PlantEye F500. This is a multispectral 3D scanner for plant phenotyping, which also provided spectral information (red, green, blue and NIR) for each point in the point clouds.

The phenotyping system was able to move to different positions, or viewpoints, around each plant, see Fig. 2 for a schematic overview. At each of these viewpoints, the mobile platform stopped and the sensor head moved in a vertical direction, starting at the plant gutter. During the upwards movement, the Phenospex F500 dual scan system generated a point cloud of the plant. The 2D images of the same plant were then collected during the downwards movement. The 2D images were collected at six height levels, from all six (I-VI) viewpoints around the plant. The 3D point clouds were collected only from viewpoints II and V, at both sides of the plant gutter, straight in front of the plant. However, since the dual scan system consisted of two scanners, each scan provided 2 viewpoints of the plant. Having multiple viewpoints reduced the number of occluded nodes. The effect of a different number of viewpoints for the 2D dataset as compared to the 3D dataset is discussed in section 4.1.

Data collection was done in 2018, between June 25th and July 5th. The 2D images were taken multiple times a day, resulting in a dataset of 9990 images. The 3D point clouds of the same plants were collected only once per day, resulting in 264 point clouds. As a reference, the internode length of all plants was measured with a measuring tape. This was done on the third, fifth and eighth day of the experiment. The internode length measurements ranged from 10 mm to 175 mm. The distribution of the internode lengths that were measured on the three days is shown in Fig. 3. The total number of internode length measurements obtained was 357. In (Boogaard et al., 2020), the mean absolute error in these measurements was estimated to be 2.7 mm, based on a comparison of multiple measurements of the same internode.

The number of nodes in a plant was always one more than the number of internodes of that plant. As the 12 plants were

[1] https://www.wiwam.be/projects/, accessed 05/08/2023.
[2] https://phenospex.com/products/plant-phenotyping/planteye-f500-multispectral-3d-laser-scanner/, accessed 05/08/2023.

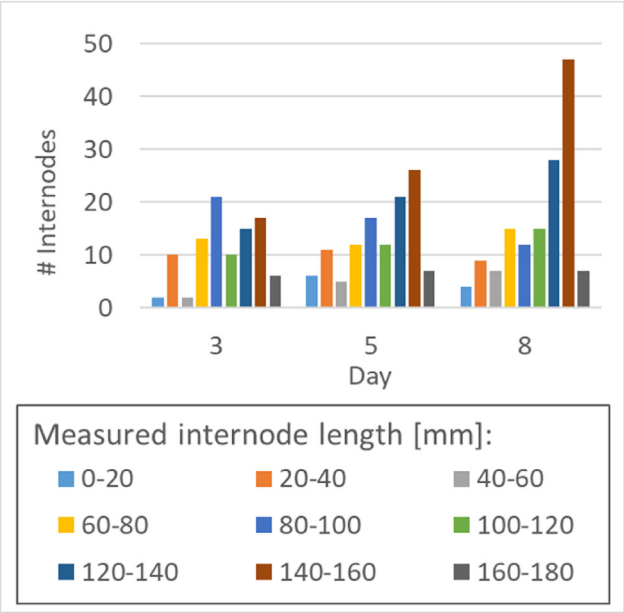

Fig. 3 – Histogram showing the distribution of internode length measured with the measuring tape on days 3, 5, and 8 of the experiment, for all plants. The bars indicate the number of internodes that fall within the bins defined in the legend of the figure.

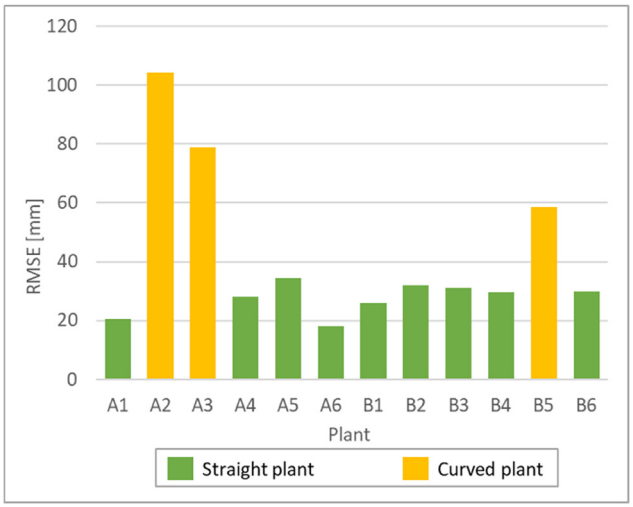

Fig. 4 – RMSE of the plant-camera distance for all plant nodes, on day 7. The curved plants A2, A3, and B5 clearly have a larger deviation from the assumed plant-camera distance than the straight plants.

measured on 3 days, the total number of nodes present in the scene was 357 internodes + (12*3) = 393 nodes.

Each plant was attached to a supporting wire to enforce vertical plant growth. However, as mentioned in the introduction, several plants still showed a curved growing pattern, making it infeasible to accurately estimate internode lengths for these plants using the 2D method. The plants for which this was the case (A2, A3 and B5) are marked in yellow in Fig. 2. These plants were identified based on visual observation in previous work by Boogaard et al. (2020). In the current work, the 3D data allowed to quantify the difference between straight and curved plants. For this purpose, the Root Mean Square Error (RMSE) was calculated, using the difference between the actual plant-camera distance and the assumed plant-camera distance as the error. The RMSE was calculated for all nodes on the last day of the experiment, as larger plants showed more curvatures due to the weight of the plant. In Fig. 4 it can be seen that, indeed, the plants that were identified by Boogaard et al. (2020) showed a higher RMSE, indicating that the plant-camera distance was not as assumed for these plants.

2.2. 2D internode length estimation

The 2D internode lengths were estimated using the method presented by Boogaard et al. (2020). For completeness, a summary of the method is added in this section. More details can be found in the mentioned paper.

The set of 2D images was used to estimate the 2D internode lengths. First, the nodes were detected in the individual images using a deep-learning based object-detection algorithm, based on YOLO-v3 (Redmon & Farhadi, 2018). Using images in which the nodes were manually labelled, the network learned to predict the location of nodes in new images. This resulted in a list of node locations per image. The node locations per image were then transformed to world coordinates, resulting in a list of detected nodes per plant.

Often, nodes were detected in multiple viewpoints. Multiple detections of the same node were combined using a clustering algorithm called affinity propagation (Frey & Dueck, 2007). The goal of the clustering algorithm was to obtain a list of node locations in which each node was present only once. The clustered list of nodes was then sorted according to location on the stem, starting with the lowest node as number 1 and counting upwards along the stem of the plant. The internode length was estimated as the Euclidean distance between two consecutive nodes. A simple linear regression model was then fitted to correct for a systematic error that was found in the estimated internode lengths.

2.3. 3D internode length estimation

In this section, the method to estimate internode lengths from the 3D point clouds is presented in three steps: i) segmentation of the point clouds into plant parts, ii) detection of nodes from the segmented point clouds, and iii) estimation of the internode lengths.

2.3.1. Point-cloud segmentation

The segmentation of the 3D point clouds into plant parts was based on the deep neural network PointNet++ (Qi et al., 2017), following the implementation presented by Boogaard et al. (2021). The neural network combines features on a per point basis, with contextual features that represent the neighbourhood of the point. To learn these contextual features, a set of points was selected from the entire point cloud. These points were chosen such that the distance between the points was maximised, meaning that all regions of the point cloud were covered. For each of these points, the neighbourhood points

were then selected as a group. For each group of points, a PointNet layer was used to learn local features within the group. Multiple of these set abstraction layers with a range of receptive fields were applied, to learn features from local to global scale. As output, the network labels every point to belong to one of the plant parts.

To train the network to learn features relevant for plant-part recognition, a training dataset was required. For the work presented in Boogaard et al. (2021), the 264 point clouds were manually segmented in stem, petiole, leaf, growing point, node, ovary, tendril, and non-plant (e.g. the plant gutter, supporting wire and stick). This was done using CloudCompare (2019). An example of a manually segmented point cloud obtained on the first and last measurement day is shown in Fig. 5.

As can be seen in Fig. 5, not all classes were equally present in the data. To improve the segmentation of the underrepresented classes, the approach presented in Boogaard et al. (2022) was followed. Due to computational limitations, it is common to divide the point cloud into smaller chunks, which are then independently processed by the network and recombined later to obtain a segmentation of the full point cloud. By sampling more chunks of underrepresented classes, the level of class balance in the training data was improved. In this way, the segmentation quality of the underrepresented classes, such as the node, significantly improved. In this work, the class-dependent sampling strategy was applied, using the default chunk size of 4096 points.

The labelled data was split in a training, validation, and test set. This was done on plant level, to prevent that point clouds of the same plant were added to multiple sets. The training set consisted of 10 plants (220 point clouds) and the validation and test set each contained the 22 point clouds of one of the remaining plants. In the current work, we needed a segmentation of all plants, therefore we repeated the data split twelve times, such that each plant was part of the test set once. Based on the twelve training sets, the network was trained twelve times, each time having a different plant in the validation and test set. The validation set was used to determine when to stop training. The weights obtained at the minimum validation loss were then used to segment the point clouds in the test set. By repeating this for all twelve test sets, a predicted segmentation of all point clouds was obtained.

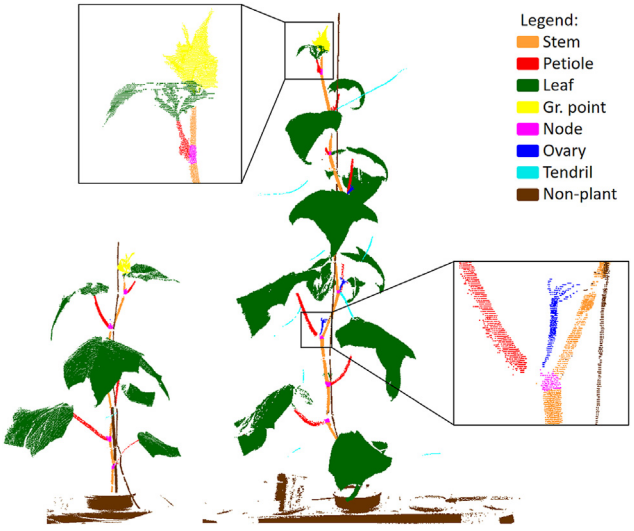

Fig. 5 – Example of a manually annotated point cloud on the first (left) and the last (right) day of data acquisition. The colours represent the classes as specified in the legend, the black squares show a zoomed in segment of the point cloud. Adapted from (Boogaard et al., 2022). (For interpretation of the references to color/colour in this figure legend, the reader is referred to the Web version of this article.)

2.3.2. *Node detection*

To reduce the level of occlusion, the point clouds obtained from viewpoints II and V, from both sides of the plant gutter, were combined. To this end, the point clouds had to be transformed from the camera coordinate system to the world coordinate system. First, the coordinate system of viewpoint II was set as the world coordinate system. Then, the point cloud obtained from viewpoint V was manually transformed in CloudCompare (2019). The origin of the point cloud was defined at the centre of the pot. A rotation of 180° around the vertical axis was applied, followed by small corrections to increase the accuracy of the registration. This procedure was done once for each plant position at the beginning of the measurement campaign. In future work, this procedure can be automated by adding a few markers to the scene. The resulting transformation matrix was stored and applied to transform all point clouds from that plant position. From the combined point clouds, all points that were predicted to be 'node' were selected.

To determine which of the selected points belonged to the same node, the node points were clustered using the HDBSCAN clustering algorithm (Campello et al., 2013). HDBSCAN is a density-based clustering algorithm that is particularly suited for finding clusters of different size and shape, in datasets that contain noise or outliers. It is not required to specify the number of clusters beforehand. The only input variable required is the minimum number of points in a cluster, called the minimum cluster size. To provide insight in the balance between number of missed nodes (false negative) and number of wrongly identified nodes (false positives), the precision, recall, and F1-score of the node detection algorithm for a range of values for the minimum cluster size are reported in section 3.2. The setting that resulted in the highest F1-score was used to obtain the node detections for the remainder of this research. The centre point of each detected node was taken as an estimate for the location of that node.

2.3.3. *Internode length estimation*

The result of the node detection step was a list of estimated node locations for each plant on each day. The node locations were sorted by increasing height, such that the first location in the list corresponded to the node closest to the pot and the last node of the list corresponded to the youngest node in the top of the plant. The internode length was then estimated as the Euclidean distance between two consecutive node locations.

Similar to the 2D case, the estimated internode lengths showed a systematic error. The main reason was that the

reference measurement was obtained following the stem of the plant, while the estimated internode length was always a straight line, resulting in an underestimation of the true internode length. Further reasons for the systematic error were not fully investigated, but possible explanations include errors in the calibration of the measurement set-up, inaccuracies in the placement of the plants and the plant gutter, and measurement errors in the collected reference measurements. For a fair comparison, the same procedure as was used for the 2D internode length estimation was used to correct for the systematic error in the 3D internode lengths. For each test plant, a simple linear regression model was fit on all other plants. That model was then used to calibrate the estimated internode lengths of the test plant.

2.4. Evaluation methods

In this section, the evaluation methods for the point-cloud segmentation, node detection, and internode-length estimation are presented.

2.4.1. Point-cloud segmentation

The evaluation of the point-cloud segmentation method was based on a comparison of the manually assigned labels and the predicted labels. A confusion matrix was used to provide insight into what classes were confused with other classes.

Furthermore, the Intersection-over-Union (IoU) for each class is presented. This metric is based on a point–wise comparison between the predicted segmentation and the manually obtained segmentation. If a point was assigned to the same class in both cases, it was considered a true positive (TP). Points that were not assigned to the same class, were considered false negatives (FN) for the manually assigned class and false positives (FP) for the predicted class. For each class c, the IoU was then calculated according to Eq. (1).

$$\mathrm{IoU_c} = \frac{\mathrm{TP_c}}{\mathrm{TP_c} + \mathrm{FP_c} + \mathrm{FN_c}}\ [-] \tag{1}$$

2.4.2. Node detection

To evaluate the detected nodes, first, the ground truth location of each node was required. All nodes that were present in the plants were given a unique ID based on the plant number, the day of measurement, and the order of the node on the stem. The nodes that were identified in the manually labelled data were then linked to the corresponding real nodes. For each manually labelled node, the centre was stored as the xyz-location of that node.

The estimated node locations from the point clouds that were segmented by the segmentation algorithm were then compared to the node locations obtained from the manually labelled data. For each of the nodes identified in the manually labelled data, the Euclidean distance to all estimated node locations was calculated. The estimated node location with the lowest distance was linked to the manually detected node and removed from the list, to make sure it was not linked to another node. If no estimated node location was found within 20 mm of the manually detected node, the node was considered not to be detected by the algorithm.

The manually detected nodes that were linked to an estimated node location were considered to be true positives (TP). The undetected nodes were considered false negatives (FN) and the estimated node locations for which no corresponding manually detected node was found were considered false positives (FP). From the number of TP, FN, and FP, the precision (P), recall (R), and F1-score (F1) of the node detection algorithm was calculated according to Eqs. (2)–(4).

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}\ [-] \tag{2}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}\ [-] \tag{3}$$

$$\mathrm{F1} = 2 \cdot \frac{\mathrm{P} \cdot \mathrm{R}}{\mathrm{P} + \mathrm{R}}\ [-] \tag{4}$$

2.4.3. Internode length estimation

The internode length measurements obtained with a measuring tape were used to evaluate the accuracy of the estimated internode lengths. Since the reference measurements were taken on the third, fifth and eighth day of the experiment, the estimated internode lengths for these days were used. Furthermore, if one or two nodes of an internode were not detected, the corresponding internode length could not be estimated and those cases were not taken into account in this evaluation. This is further discussed in section 4.3. The number of estimated internode lengths that was used for the evaluation is reported in section 3.4.

The difference between the estimated internode length, $\hat{s}_i$, and the manually measured internode length, s_i, was called the error e_i, for the internode length between node i and node $i+1$, according to Eq. (5).

$$e_i = \hat{s}_i - s_i\ [\mathrm{mm}] \tag{5}$$

To test the significance of differences in the errors between the 2D internode lengths and the 3D internode lengths, and between the curved plants and the straight plants, a Mann–Whitney U test was used.

3. Results

The results of this paper are presented in this section. The segmentation performance is presented in section 3.1, showing how well the plant parts were identified in the point clouds. In section 3.2, the results of the clustering algorithm are given for different values of the minimum cluster size. The visibility and detection of the nodes is then presented in section 3.3. Finally, in section 3.4, the results of the estimated internode lengths are presented and compared to the internode lengths obtained from the 2D data. Intermediate results from the 2D method are not presented, these results can be found in (Boogaard et al., 2020).

3.1. Point-cloud segmentation

The point clouds of the twelve test sets were segmented by the corresponding twelve trained models. The predicted labels were then compared to the manually assigned labels. The

resulting confusion matrix is shown in Table 1. Each row corresponds to the points that were manually assigned to that class and shows how these points were labelled by the network. The correctly predicted points are shown on the diagonal. The best performance was obtained for the classes 'leaf' (98.8%) and 'non-plant' (98.1%). Performance decreased for 'stem', 'growing point', 'petiole', 'tendril', and 'ovary'. The lowest percentage of correctly predicted points was observed for the class 'node' (43.0%). The majority of incorrectly classified 'node' points was predicted to be 'stem'.

The mean IoU-values per class are presented in Fig. 6, including the 95%-confidence interval on the mean. Again, the classes 'leaf' and 'non-plant' showed the best performance, while the lowest mean IoU-value was observed for the 'node'. The small confidence intervals indicate a consistent performance over the point clouds in the test sets.

3.2. *Clustering algorithm*

The points that were predicted to be 'node' were clustered by the HDBSCAN-algorithm. The effect of different values for the input parameter 'minimum cluster size' was tested and the results are shown in Fig. 7. The highest precision (0.93) was found for a minimum cluster size of 50 points. For larger values, the number of correctly detected nodes dropped, since smaller nodes were either discarded or merged with other nodes in their neighbourhood. For smaller minimum cluster sizes, the number of correctly detected nodes kept increasing, however, the precision was lower due to a higher number of false positive detections. These false positive detections do not have an effect on recall, and therefore, the highest recall (0.85) was found for a minimum cluster size of 2 points. The decrease in recall for larger values of the minimum cluster size is due to an increasing number of false negatives.

The maximum value of the F1-score (0.87) was observed for a minimum cluster size of 20 points. Therefore, this setting was used for the remainder of this research.

3.3. *Node detection*

As mentioned in section 2.1, the total number of nodes was 393. However, not all of the nodes that were present in the plants, were also found in the data. Based on the manually labelled point clouds, the number of nodes that was identified by the annotator was 293. This means that 100 of the real nodes, were not identified by the annotator in the collected point clouds. Further analysis of these missing nodes showed that 54 of them concerned the two youngest nodes of the plant. During manual annotation of the point clouds, the top of the plant was labelled as 'growing point'. It turned out that this class overlaps with the youngest nodes, which were therefore not labelled and also not learned by the segmentation algorithm. The remaining 46 nodes that were not identified in the point clouds were mostly occluded by leaves.

Based on the node points that were predicted by the segmentation algorithm, the clustering algorithm found 257 clusters of node points. After linking these clusters to the manually annotated nodes, 238 clusters were identified as correctly predicted nodes (true positive), meaning that the remaining 19 clusters were false positives. Furthermore, this means that 293−238 = 55 nodes were not detected, or false negatives. Based on this data, the precision of the node detection algorithm was 0.93 (238/257) and the recall was 0.81 (238/293), leading to an F1-score of 0.87. In comparison, the precision and recall of the node detection step for the 2D method were 0.95 and 0.92 respectively, leading to an F1-score of 0.93 (Boogaard et al., 2020).

The results of the node detection are schematically summarised in Fig. 8. Note that, even while only 43.0% of the 'node' points in the point clouds were correctly segmented (see Table 1), as long as part of the points of a node were found, the clustering algorithm could detect this node.

3.4. *Internode length estimation*

The estimated internode lengths were evaluated for internodes for which both nodes were correctly detected, as explained in section 2.4.3. The number of internodes that was present in the plants, based on the manual measurements, was 357. The 2D method was able to correctly detect both nodes for 346 of these internodes. In the 3D point clouds, the annotator identified both nodes for 230 internodes. The 3D node detection algorithm detected both nodes for 171 internodes. The results presented in this section are based only on correctly detected node pairs, other node pairs were manually removed as explained in section 2.3.3. This is further discussed in section 4.3.

Table 1 – Confusion matrix of the trained networks. Each row corresponds to the points that were manually labelled as that class, showing how the network predicted those points. The values on the diagonal are correct predictions and correspond to the recall.

Confusion matrix (percentages)		Predicted labels							
		Stem	Petiole	Leaf	Gr. point	Node	Ovary	Tendril	Non-plant
Manually assigned labels	Stem	90.2	2.1	1.3	1.1	2.9	0.3	0.4	1.6
	Petiole	6.9	82.5	3.9	1.0	2.8	1.6	0.9	0.5
	Leaf	0.1	0.1	98.8	0.6	0.0	0.0	0.1	0.3
	Gr. point	1.6	0.3	8.1	88.4	0.0	0.0	0.5	1.2
	Node	44.0	7.0	1.2	1.2	43.0	1.1	0.8	1.7
	Ovary	9.0	11.9	2.9	0.3	4.9	67.6	1.8	1.5
	Tendril	6.4	3.2	8.8	1.2	1.5	1.2	72.6	5.1
	Non-plant	0.9	0.1	0.4	0.1	0.0	0.0	0.4	98.1

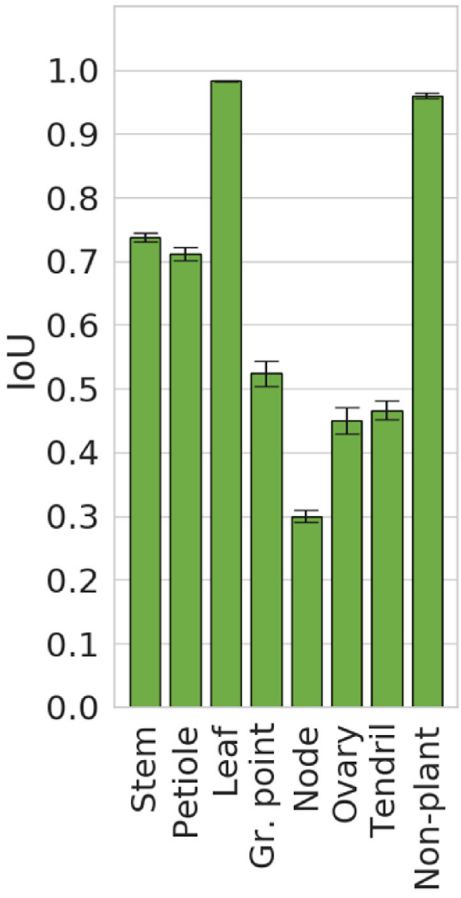

Fig. 6 – IoU-values based on the predicted segmentation of the test sets. The bars indicate the mean IoU-values and the error bars show the 95% confidence interval on the mean.

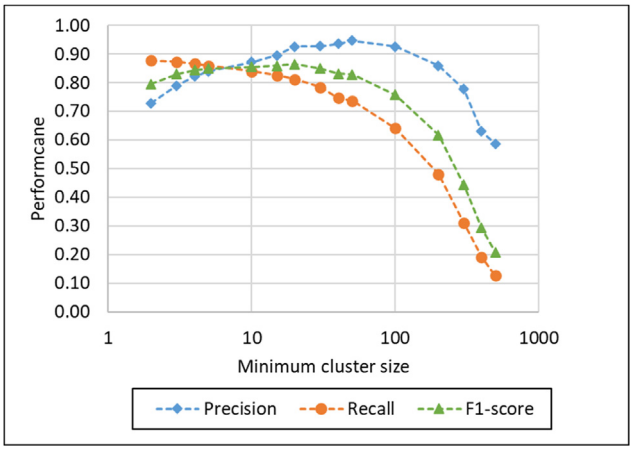

Fig. 7 – Precision (blue diamonds), recall (orange circles) and F1-score (green triangles) of the clustering algorithm for different values of the minimum cluster size used by the HDBSCAN-algorithm. Note that the x-axis, showing the minimum cluster size, has a logarithmic scale for readability of the figure. The highest F1-score was observed for a minimum cluster size of 20, this value was used in the remainder of this research. (For interpretation of the references to color/colour in this figure legend, the reader is referred to the Web version of this article.)

The manually measured internode lengths were plotted against the estimated internode lengths from the 2D images and from the 3D point clouds, as shown in Fig. 9. First, what is clearly visible, is that the 3D internode lengths were more accurate than the 2D internode lengths. Furthermore, there was no visible difference between the straight plants and the curved plants in the results of the 3D method, while the curved plants clearly have a larger error in the results of the 2D method.

In the lower-left corner of the two plots, it can be seen that the 2D method was able to estimate shorter internode lengths than the 3D method. As mentioned in section 2.1, the shortest internode length that was measured with the measuring tape was 10 mm. However, the shortest internode for which the 3D method detected both nodes and thus could estimate the internode length, was manually measured to be 45 mm. So, despite the higher accuracy of the 3D method, short internodes were not measured. The main reason for this is that the nodes in the top of the plant, corresponding to the shorter internodes, were overlapping with the class 'growing point', as explained in section 3.3.

A boxplot summarising the absolute errors of the internode length estimates, for straight and curved plants, and for 2D images and 3D point clouds is shown in Fig. 10. To test if the internode lengths estimated for the curved plants were significantly different from the internode lengths estimated for the straight plants, a Mann–Whitney U test was used. For the 2D images, the null hypothesis was rejected ($p < 0.001$), indicating that internode lengths estimated by the 2D method for the straight plants were significantly more accurate than for the curved plants. Also, for the 3D method, the absolute errors for the straight plants were compared to the absolute errors for the curved plants. In this case, the null hypothesis was accepted ($p = 0.29$), meaning that the absolute errors for the straight plants were not significantly different from the errors for the curved plants. In other words, indeed, the 3D method is able to measure the curved plants with the same accuracy as the straight plants.

Besides the comparison between curved and straight plants, it was tested if the internode lengths estimated from the 3D point clouds were more accurate than the internode lengths from the 2D images, as Fig. 9 suggested. It was found that the absolute errors of the 3D method for straight plants (median: 4.1 mm, mean: 5.2 mm) were significantly lower ($p < 0.001$) than the absolute errors of the 2D method for straight plants (median: 6.33 mm, mean: 7.7 mm). Furthermore, also the absolute errors of the 3D method for curved plants (median: 4.0 mm, mean: 4.7 mm) were found to be significantly lower ($p < 0.001$) than the absolute errors of the 2D method for straight plants.

4. Discussion & recommendations

The results presented in this paper clearly show that the internode lengths estimated from the 3D point clouds were

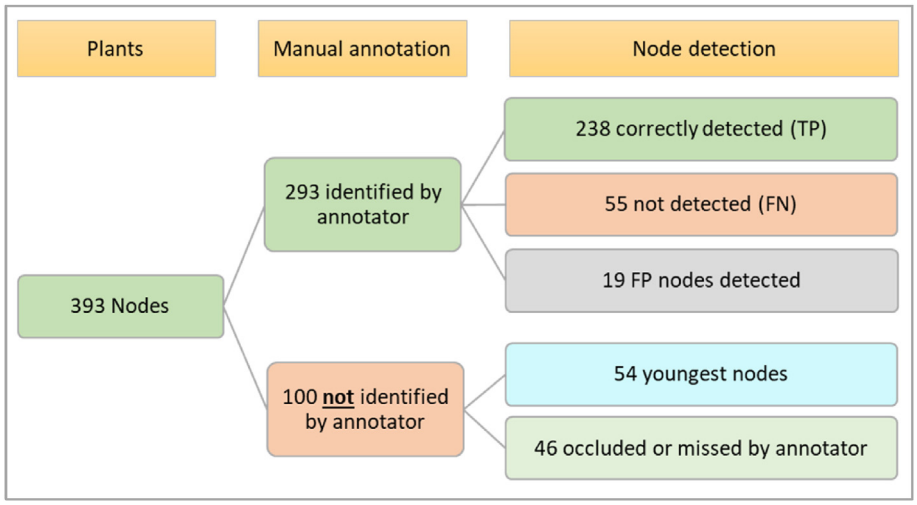

Fig. 8 – Classification of the nodes that were present in the plants, divided over identified and not identified by the annotator of the point clouds. For the identified nodes, the number of detected nodes for the 3D method is reported.

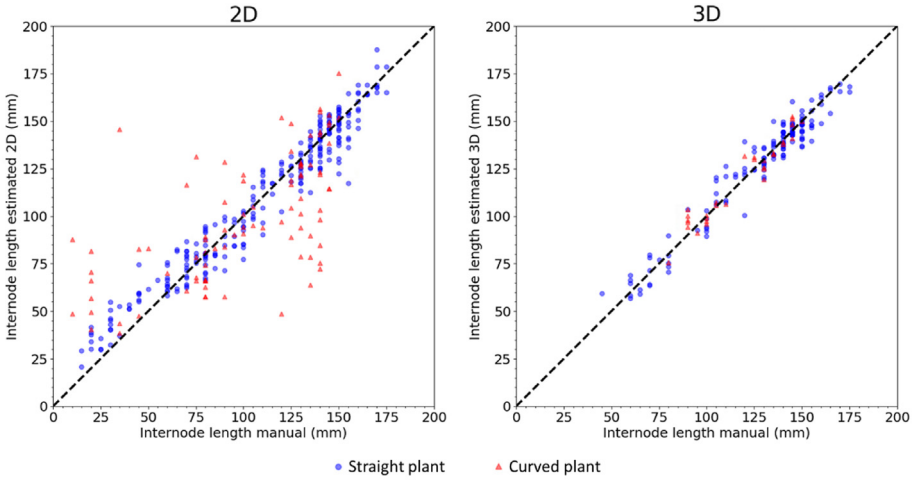

Fig. 9 – The manually measured internode length (x-axis) plotted against the estimated internode length (y-axis). The left plot shows the results based on the 2D method, the right plot shows the results of the current work based on the 3D point clouds. The nodes of the straight plants are shown as blue circles, while the nodes of the curved plants are shown as red triangles. The dashed line shows the function x = y, as an indication of where the estimated internode length was equal to the reference measurement. (For interpretation of the references to color/colour in this figure legend, the reader is referred to the Web version of this article.)

more accurate than the internode lengths estimated from the 2D images. However, as mentioned in section 3.4, the number of detected nodes and thus the number of internode lengths that could be estimated was higher for the 2D method. In this section, these results will be further discussed, starting with the node visibility in section 4.1. The node detection is discussed in section 4.2. The internode length estimations are discussed in section 4.3, and the final recommendations are given in section 4.5.

4.1. *Node visibility*

The level of occlusion, or the total number of nodes that was present in the acquired data, depended on the number of viewpoints used. For the 2D images, it was shown that the percentage of visible nodes increased from 91.9% to 99.3%, when all six viewpoints around the plant were used instead of only viewpoint II and V (Boogaard et al., 2020). In comparison, the percentage of visible nodes in the 3D point clouds was 74.6% (293/393). Out of the 100 nodes that were not visible, 46 nodes were occluded by leaves or other plant parts. Based on the increased visibility rate observed in the 2D data, it is likely that also in the 3D data, more nodes would be visible when adding more viewpoints. An alternative would be to use an active vision approach (Burusa et al., 2022; Wu et al., 2019). In such a system, the camera could be mounted on a robot arm, providing more flexibility to use additional viewpoints. In combination with an online node recognition method, the

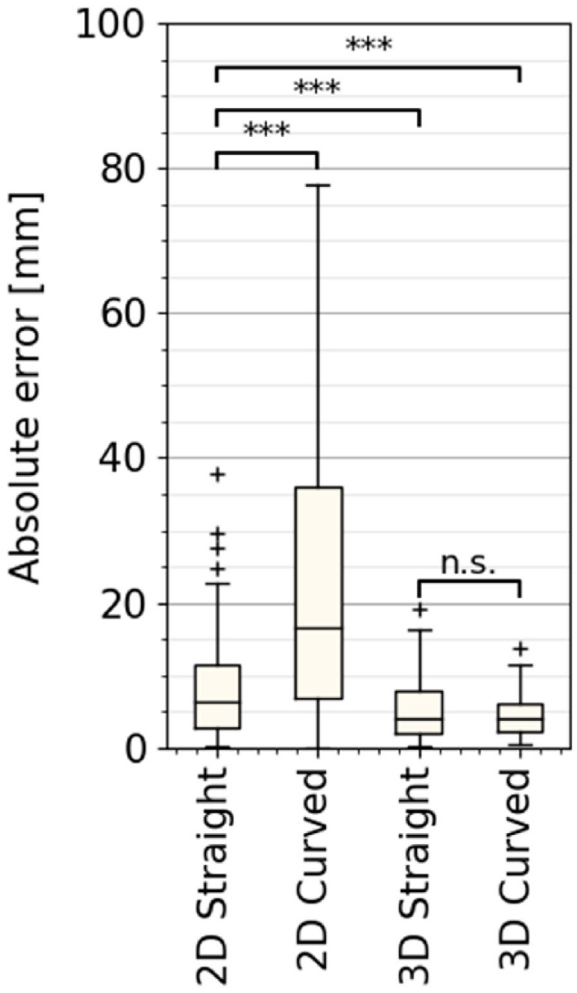

Fig. 10 – Absolute error (mm) of the predicted internode lengths based on 2D and 3D data, for straight and curved plants. The boxes show the lower and upper quartile, the whiskers indicate the highest and lowest absolute error within 1.5 times the inter-quartile range. Values outside this range were considered outliers and are shown as '+'. The level of significance above the arches is indicated as 'n.s.', meaning no significant difference was observed, or as '*', meaning a significant difference was observed with $p < 0.001$.**

system could be optimised to add new viewpoints until as many nodes as possible are included in the data.

The remaining 54 nodes that were not identified by the annotator were labelled as 'growing point'. The height of the annotated growing point was in the range of 50–70 mm, an example can be seen in Fig. 5. When collecting the reference measurements, nodes were measured if they were longer than 10 mm, meaning that often the 1 or 2 youngest nodes were hidden inside the growing point. However, the resolution of the point clouds was not sufficient to identify these nodes in the point cloud. It could be tested if other sensors or a combination with the 2D images, which had a higher resolution, could improve the detection of the youngest nodes. However, this also leads to the question whether it is relevant to measure internode length in such an early stage. This goes beyond the scope of the current work, but the specific requirements of automated internode length measurements should be discussed in more detail with plant scientists in future work. The results presented in this paper contribute to this discussion, by providing benchmarks of what can be expected when using different approaches.

4.2. *Node detection*

Besides the higher visibility rate of nodes in the 2D dataset, also the detection rate was higher. The F1-score for node detection from the 2D images was 0.93, while the F1-score for the detection of nodes from the 3D point clouds was 0.87. Both the number of false positives as well as the number of false negatives has a negative effect on the F1-score, which will be discussed in the next two paragraphs.

The total number of false positive node detections in the 3D data was 19. Further inspection showed that 15 of the false positives could be explained due to other plant parts that were predicted as nodes. In the remaining 4 cases, the node points of two nodes that were close to each other, were combined into one cluster. The centre of such a cluster, which was used as the estimated node location, was in the middle between the two nodes. Therefore, the distance to either of these two nodes was too large, such that the cluster was not linked to the manually detected nodes and was considered a false positive. Also, since this link was not made, the two manually detected nodes remained undetected and were both considered a false negative.

The total number of false negative detections in the 3D data was 55. These nodes were mainly undetected because the node points were predicted to be stem or petiole, as can be seen in Table 1. The low IoU for the class 'node' (0.3) also indicates that the segmentation quality could be improved. One approach would be to increase the amount of training data further, aiming to better represent the variation in the data. However, obtaining high quality training datasets for point cloud segmentation, especially in the plant domain, is still challenging (Boogaard et al., 2021; Dutagaci et al., 2020; Turgut et al., 2022). Alternatively, attempts to focus on node detection only, instead of having a full segmentation of the point cloud, might improve the detection rate. However, when looking at the broader perspective of measuring plant architecture, a properly segmented point cloud is a valuable asset, since it opens up possibilities to measure other plant traits.

4.3. *Internode length estimation*

In the results section, it was shown that the proposed method to estimate 3D internode length improved the internode length estimates for curved plants and outperformed the previous 2D method in terms of accuracy also for the straight plants. However, this is based only on the internodes for which the two corresponding nodes were detected. If a specific node was not detected, the internode length below as well as the internode length above the undetected node could not be estimated.

In the current paper, both the evaluation of whether a detected node was correct, as well as the check whether two nodes were belonging to the same internode, was done based

on the manually annotated data. While this is a valid approach to answer the research questions that were dealt with in this paper, it hampers the practical applicability of the method. Therefore, besides improving the node detection step, further research into automated detection of missing nodes is recommended. One approach could be to use the segmented stem elements to determine if two detected nodes are connected. Using this information would also allow to measure the actual internode length following the contours of the stem, instead of the Euclidean distance between two nodes.

4.4. *Generalisation*

The current work presents a comparison of a 2D and a 3D method for estimating internode length in cucumber. The dataset that was used is limited in the sense that only one cucumber variety was used, grown in one environment and assessed using one data acquisition system. This dataset does not allow to test generalisation of the presented methods to other varieties, crops, and environments. However, all methods were set-up in a generic way, and the only crop-specific part was the training data. Therefore, we expect that the methods should generalise well, although retraining might be required.

A possible limitation of our methods was in the spacing of the plants. In the current dataset, no overlap between plants was present. It can be expected that performance drops when moving to more complex and cluttered environments. Introducing overlap between plants also adds the task of allocating detected plant parts to individual plants and keeping track of this over time. It is recommended that this task is subject of future research.

4.5. *Final recommendations*

Both the 2D and the 3D method have their advantages and disadvantages. One of the main limitations of the 2D method, was that not all plants could be accurately measured, especially if the plants showed a large variation in plant-camera distance. For the 3D method, one of the main limitations was that not all nodes were visible and detected, meaning that not all internode lengths could be estimated. To overcome the limitations of the 2D and the 3D method, it is recommended to combine both datasets. It is expected that a combined approach that uses the high resolution of the 2D images for accurate detection and the spatial information contained in the 3D point clouds to correct for variations in plant-camera distance can solve the limitations indicated above.

Still, already in the current form, the point clouds and the developed methods deliver quantitative information about the plant architecture. Furthermore, the point clouds can be segmented, and this work could serve as an example on how to measure traits from the segmented point cloud. Future research could focus on data acquisition systems that reduce the level of occlusion, point cloud segmentation methods that improve the segmentation and can handle a higher level of phenotypic variation, and on methods to extract additional phenotypic measurements from the segmented point clouds.

5. Conclusion

The aim of this paper was to demonstrate the added value of 3D computer-vision based methods for measuring plant architecture. Previous work focussed on the semantic segmentation of the 3D point clouds. The current work extends these methods by using the segmented point clouds to extract actual plant measurements. It was demonstrated that from the segmented point clouds, nodes can be detected and internode lengths can be estimated.

The first hypothesis that was tested, was that *"using 3D point clouds instead of 2D images allows to estimate internode lengths of curved plants with the same accuracy as obtained for straight plants"*. The results showed that based on the 3D method presented in this work, there was indeed no difference in accuracy between internode lengths estimated for the curved and the straight plants. Moreover, the hypothesis was that *"the error of the internode lengths estimated from the 3D point clouds is smaller than the error of the internode lengths estimated from the 2D images"*. Indeed, the results showed that the error obtained with the proposed 3D method in curved plants and in straight plants was significantly lower than the error obtained with the 2D method in straight plants.

A limitation of the 3D method was that not all nodes were detected due to occlusion and a lower resolution of the point clouds as compared to the 2D images. It was suggested to explore the possibilities of a combined approach, in which the advantages of both the 2D images and the 3D point clouds could be utilised. Still, in this paper, it was demonstrated that computer-vision based measurements of plant architecture in general, and more specifically of internode length, greatly benefit from the availability of 3D data.

Funding

Frans Boogaard gratefully acknowledges Rijk Zwaan for the funding of his position as PhD-candidate. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Boogaard, F. P., Rongen, K. S. A. H., & Kootstra, G. W. (2020). Robust node detection and tracking in fruit-vegetable crops using deep learning and multi-view imaging. *Biosystems Engineering*, 192, 117–132. https://doi.org/10.1016/j.biosystemseng.2020.01.023

Boogaard, F. P., van Henten, E. J., & Kootstra, G. (2021). Boosting plant-part segmentation of cucumber plants by enriching incomplete 3D point clouds with spectral data. *Biosystems*

Engineering, 211, 167–182. https://doi.org/10.1016/j.biosystemseng.2021.09.004

Boogaard, F. P., van Henten, E. J., & Kootstra, G. (2022). Improved point-cloud segmentation for plant phenotyping through class-dependent sampling of training data to battle class imbalance. *Frontiers in Plant Science, 13*(March). https://doi.org/10.3389/fpls.2022.838190

Burusa, A. K., van Henten, E. J., & Kootstra, G. (2022). *Attention-driven active vision for efficient reconstruction of plants and targeted plant parts*. http://arxiv.org/abs/2206.10274.

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Lecture Notes in Computer Science, 7819 LNAI*(PART 2), 160–172. https://doi.org/10.1007/978-3-642-37456-2_14

CloudCompare. (2019). *CloudCompare* (version 2.10.2) (2.10.2) http://www.cloudcompare.org/.

Costa, C., Schurr, U., Loreto, F., Menesatti, P., & Carpentier, S. (2019). Plant phenotyping research trends, a science mapping approach. *Frontiers in Plant Science, 9*(January), 1–11. https://doi.org/10.3389/fpls.2018.01933

Dutagaci, H., Rasti, P., Galopin, G., & Rousseau, D. (2020). ROSE-X: An annotated data set for evaluation of 3D plant organ segmentation methods. *Plant Methods, 16*(1), 1–14. https://doi.org/10.1186/s13007-020-00573-w

Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points (supporting online material). *Science, 315*(5814), 972–976. https://doi.org/10.1126/science.1136800

Gehan, M. A., & Kellogg, E. A. (2017). High-throughput phenotyping. *American Journal of Botany, 104*(4), 505–508. https://doi.org/10.3732/ajb.1700044

IMPERX. (2018). *IMPERX industrial cameras & imaging systems*. https://www.imperx.com/ccd-cameras/b4820/.

Litvin, A. G., Van Iersel, M. W., & Malladi, A. (2016). Drought stress reduces stem elongation and alters gibberellin-related gene expression during vegetative growth of tomato. *Journal of the American Society for Horticultural Science, 141*(6), 591–597. https://doi.org/10.21273/JASHS03913-16

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

Najla, S., Vercambre, G., Pagès, L., Grasselly, D., Gautier, H., & Génard, M. (2009). Tomato plant architecture as affected by salinity: Descriptive analysis and integration in a 3-D simulation model. *Botany, 87*(10), 893–904. https://doi.org/10.1139/B09-061

Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). *PointNet++: Deep hierarchical feature learning on point sets in a metric space*. http://arxiv.org/abs/1706.02413.

Redmon, J., & Farhadi, A. (2018). *YOLOv3: An incremental improvement*. https://doi.org/10.1109/CVPR.2017.690

Reinhardt, D., & Kuhlemeier, C. (2002). Plant architecture. *EMBO Reports, 3*(9), 846–851. https://doi.org/10.1093/embo-reports/kvf177

Sibomana, I. C., Aguyoh, J. N., & Opiyo, A. M. (2013). Water stress affects growth and yield of container grown tomato (Lycopersicon esculentum Mill) plants. *Global Journal of Bio-Science and Biotechnology, 2*(4), 461–466. http://www.scienceandnature.org/GJBB_Vol2(4)2013/GJBB-V2(4)2013-1.pdf.

Tripodi, P., Nicastro, N., & Pane, C. (2022). Digital applications and artificial intelligence in agriculture toward next-generation plant phenotyping. *Crop & Pasture Science*. https://doi.org/10.1071/CP21387

Turgut, K., Dutagaci, H., Galopin, G., & Rousseau, D. (2022). Segmentation of structural parts of rosebush plants with 3D point-based deep learning methods. *Plant Methods, 18*(1), 20. https://doi.org/10.1186/s13007-022-00857-3

Wu, C., Zeng, R., Pan, J., Wang, C. C. L., & Liu, Y. J. (2019). Plant phenotyping by deep-learning-based planner for multi-robots. *IEEE Robotics and Automation Letters, 4*(4), 3113–3120. https://doi.org/10.1109/LRA.2019.2924125