Three-dimensional leaf edge reconstruction using a combination of two- and three-dimensional phenotyping approaches

Hidekazu Murata
Kyushu University, Fukuoka

Koji Noshita (noshita@morphometrics.jp)
Kyushu University, Fukuoka

Research Article

Keywords: Phenomics, Leaf, Image analysis, Photogrammetry, Machine learning, Soybean

Posted Date: September 19th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3347414/v1

License: ☑️ ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background: The physiological functions of plants are carried out by leaves, which are important organs. The morphological traits of leaves serve multiple functional requirements and demands of plants. Traditional techniques for quantifying leaf morphology rely largely on two-dimensional (2D) methods, resulting in a limited understanding of the three-dimensional (3D) functionalities of leaves. Notably, recent advancements in surveying technologies have improved 3D data acquisition processes. However, there are still challenges in producing accurate 3D-representations of leaf morphologies, particularly leaf edges. Therefore, in this study, we propose a method for reconstructing 3D leaf edges using a combination of 2D image instance segmentation and curve-based 3D reconstruction.

Results: The proposed method reconstructed 3D leaf edges from multi-view images based on deep neural network-based instance segmentation for 2D edge detection, SfM for estimating camera positions and orientations, leaf correspondence identification for matching leaves among multi-view images, curve-based 3D reconstruction for estimating leaf edges as 3D curve fragments, and B-spline curve fitting for integrating curve fragments into a 3D leaf edge. The method was demonstrated on both virtual and actual plant leaves. On the virtually generated leaves, we evaluated the accuracy of the 3D reconstruction by calculating standardized Fréchet distance, which reveals that small leaves and high camera noise pose greater challenges to reconstruction. To balance the number and precision of 3D curve fragments, we proposed guidelines for setting the threshold for how only reliable curve fragments are reconstructed based on simulated data. These guidelines suggested that the threshold becomes lower with greater occlusions, larger leaf size, and camera positional error greater than a certain level. We also found the number of images does not affect the optimal threshold except in very few cases. Moreover, the proposed method succeeded in reconstructing holes in the leaf when the number of holes is three or less.

Conclusions: In this study, a nondestructive method for 3D leaf edge reconstruction was developed to address the 3D morphological properties of plants, which have been challenging to evaluate quantitatively. It is a promising way to capture whole plant architecture by combining 2D and 3D phenotyping approaches adapted to the target anatomical structures.

Background

Leaves are highly important organs for plants since they are the sites of fundamental physiological processes, including photosynthesis, transpiration, and respiration. The phenotypic diversity of leaves underlies the various functional demands associated with their habitats [1–4]. Furthermore, their morphological properties are essential in balancing the multiple functional demands of individual plants and canopies [5–7], such as light interception [8, 9], heat transfer [10, 11], hydraulic conductivity [12, 13], mechanical constraints [14, 15], and growth efficiency [16, 17]. Quantifying the morphological traits of leaves provides a quantitative understanding of the relationships between the morphological traits and genetics of plants, morphogenesis, and environmental conditions, providing valuable insights into plant growth and development, improving crop yields, and enhancing plant productivity.

Leaves have complex three-dimensional (3D) shapes. Despite this, traditional measurement, quantification, and evaluation techniques rely on two-dimensional (2D) methods because they are simple and more feasible to use, especially considering existing technical limitations. In many cases, botanical specimens are preserved two-dimensionally [18] and undergo morphological changes upon drying [19, 20]. Quantitative evaluations are based on 2D imaging (e.g., flatbed scanners) and image analysis (e.g., [21, 22]). Leaves exhibit a wide range of patterns in 3D shapes [23], and their functionality is highly dependent on their configuration in the 3D space [24–27]. Moreover, some shapes cannot be adequately projected two-dimensionally (e.g., twisted leaves of Codiaeum variegatum 'Spiralé'). Consequently, many leaf characteristics have not been appropriately evaluated through 2D methods, inspiring interest in 3D evaluations.

High-resolution 3D morphological data can be acquired efficiently and cost-effectively using light detection and ranging (LiDAR) sensors, depth cameras, and photogrammetry techniques [28–30]. A pipeline utilizing structure from motion (SfM) and multi-view stereo (MVS), which reconstructs a 3D surface as point cloud data from a series of 2D images captured from different angles, has been implemented in several libraries and software products (e.g., [31, 32]). Several devices and techniques for acquiring the structures of plants in 3D have been developed to facilitate 3D evaluation in plant phenotyping studies [33, 34]. However, point cloud data produced by point-based 3D reconstruction methods, such as the commonly used SfM/MVS pipeline, may not be ideal for
representing 3D leaf morphologies because of unclear leaf edges [35] and uncertainties regarding whether the holes in point cloud data are actually real or results of reconstruction errors [36]. Point-cloud data reconstructed using the point-based reconstruction method often include points representing both leaves and artifacts owing to the keypoints detected on the background (Fig. 1a). The holes in the output point-cloud data comprise reconstruction deficiencies and actual holes; it is difficult to distinguish between them solely based on point-cloud data (Fig. 1b, c). It is preferable to establish phenotyping methods that enable the direct estimation of leaf edges.

In this study, we proposed a method to reconstruct leaf edges from multi-view images using deep learning-based instance segmentation for 2D edge detection (Fig. 2a, b), SFM for estimating camera positions and orientations (Fig. 2c), leaf correspondence identification for matching leaves among multi-view images (Fig. 2d), curve-based 3D reconstruction for estimating leaf edges as curve fragments in 3D spaces (Fig. 2e), and B-spline curve fitting for integrating curved fragments into 3D leaf outlines (Fig. 2f). The applicability and limitations of the proposed method were examined using both simulated data and actual multi-view images of soybean plants. Our analysis revealed that leaf size, error in camera parameter estimation, and mask estimation error had significantly impacted accuracy. The proposed method is expected to be a valuable tool for clarifying the morphological characteristics of 3D leaf edges, which are difficult to quantitatively evaluate.

Results

3D leaf edge reconstruction on virtually generated leaf models

The key idea of the proposed method is directly estimating 3D leaf edges using curve-based 3D reconstruction. In this study, we used a 3D curve sketch [37] that reconstructed a set of 3D curve fragments from the 2D edges of a target object in multi-view images (Fig. 3 and Methods). Based on the principle of epipolar geometry, candidates for corresponding edge pairs were identified along the epilines (Fig. 3a). All candidate edge pairs, which were called pair hypotheses, were reconstructed in 3D space; and these 3D curve fragments were reprojected onto other images to evaluate how closely the reconstructed curve generated the true projection (Fig. 3b). If the reprojected curve was sufficiently close to the edges in the reprojected image, it indicated that the pair hypothesis had been supported by the reprojected image. Only 3D curve fragments supported by a number of images greater than the support threshold \( \tau_s \) were reconstructed (Fig. 3c). The proposed method was first demonstrated on virtual data generated from single and multiple leaves created using Blender [38] (see Methods for the details).

On single virtual leaves, true mask images and camera parameters are known. Based on this assumption, 3D leaf edges were reconstructed by extracting the 2D leaf edges from true mask images and adopting a 3D curve sketch (Fig. 4, upper row); the reconstructed leaf edges appeared along the edges. Notably, the support threshold \( \tau_s \) strongly affected the performance of curve-based reconstruction; low \( \tau_s \) values resulted in highly inaccurate 3D curve fragments, and high values resulted in 3D curve fragments that did not completely cover the leaf edges (Additional file 1: Fig. S1). Details regarding \( \tau_s \) adjustment are discussed later (see the subsection "Guidelines for setting support thresholds in 3D edge reconstruction with occlusions").

Regarding the 3D edges of multiple virtual leaves of a single plant, they were reconstructed after identifying the correspondences between individual leaves across the mask images, resembling reconstruction in the single-leaf case in all aspects except for considering the influence of occlusion (Additional file 1: Fig. S2a). However, the correspondence of leaves among mask images is nontrivial in actual multi-view images because the mask image is estimated for each individual image. Thus, we precisely estimated the 3D leaf edges of multiple leaves in a single scene by incorporating a leaf correspondence identification step that prevented the generation of pair hypotheses between non-corresponding leaves across views (Fig. 4, lower row, Methods, and Additional file 2). In the absence of leaf correspondence identification, the number of reconstructed curve fragments decreased, and the vertical reconstruction error increased (Additional file 1: Fig. S2b). In the leaf correspondence identification step, the point cloud data obtained from SFM (Fig. 2c) was clustered into each leaf, and each cluster corresponding to a single leaf was reprojected onto the mask images (Fig. 5a). To preclude the leakage of points in backside into the front during reprojection, hidden point removal [39] was applied from each view (Fig. 5b). Leaf correspondences were identified by counting the number of reprojected points belonging to each cluster in each image (Fig. 5c).

Comparing the accuracy of 3D leaf edge reconstruction under different conditions
We evaluated the accuracy of the 3D leaf edge reconstruction method for different leaf areas, image numbers, occlusion levels, and noise levels. This evaluation was performed on the simulated multiple leaf data using the standardized Fréchet distance \( [40] \) (SFD) divided by the square root of the leaf area. The SFD was calculated for three individual plants with 8 different-sized leaves \( (312\text{mm}^2 \leq A \leq 3,366\text{mm}^2) \) in several simulation scenarios, including different levels of occlusion (no, thin, and thick pillars), different numbers of multi-view images (32, 64, and 128 images), and different degrees of positional noise affecting the camera parameters \( (\sigma = 0, 1, \text{and} 3 \text{mm}) \) (Fig. 6).

The SFD decreased with the increase in leaf area; the small leaves were more challenging to reconstruct than the larger leaves were. Small leaves had larger curvatures even if they had the same shapes, making it difficult for the 3D curve sketch to reconstruct the correct curve fragments because 2D curve fragments had been frequently generated through splitting by a tangential epipolar line (see Methods for details). The SFD increased with increases in the degree of noise at the camera positions. Although a less accurate camera extrinsic parameter estimation would increase the SFD, the effect might be limited under low noise. However, the SFD was less sensitive to the number of images and level of occlusion, considering that even if a leaf edge was obscured in an image, it could be complemented if it had appeared in other images \([41]\).

Generation of mask images from actual multi-view images using Mask R-CNN

To generate a mask image for each leaf from multi-view images of actual plants, we used Mask R-CNN \([42]\), which is a deep neural network (DNN) model used for instance segmentation. The performance of the model was evaluated on 1039 images of four individuals with group four-fold cross-validation; the images belonging to one individual were excluded as test images, and the other images were randomly divided into training (80%) and validation (20%) datasets (Fig. 7a and Methods). The model weights were adopted at epoch 8000, and the validation loss did not improve thereafter on the learning curve until epoch 10000 (Additional file 1: Fig. S3). Individual leaf masks were generated using the trained model (Fig. 7b). The average precision (AP) values of the test data are listed in Table 1. Regarding the values, AP was 49.8, and AP large (APl) was 76.9, indicating that inference had been successful in a large region, resembling the trend in a previous study on generic object recognition \([42]\). On the other hand, AP middle (APm) and AP small (APS) were smaller than APl, suggesting that generating masks for small leaves had been challenging. The model weights trained on all images of the four individuals were used for the subsequent analysis.

Guidelines for setting support thresholds in 3D edge reconstruction

To obtain accurate 3D leaf edges, the support threshold \( (\tau_t) \) should be set appropriately to balance the trade-off between the number and precision of the reconstructed 3D curve fragments. We attempted to propose optimal support thresholds against occlusion indices (OIs) based on simulated virtual leaves by evaluating the precision-recall (PR) curve of the reconstructed 3D edges (i.e., the optimal support threshold was defined as the highest support threshold that resulted in the highest recall when the precision had exceeded 0.99) (Fig. 8 and Methods). In this study, the occlusion index of a target leaf was defined based on the sparse point cloud data of the target, as follows:

\[
OI = \frac{1}{m} \sum_{i=1}^{m} \left(1 - \frac{n_i}{n}\right)
\]

where \( m \) is the number of images; \( n \) is the number of points of a target instance; and \( n_i \) is the number of points of a target instance reprojected onto the \( i \)-th image; and \( OI \) is the occlusion index, where \( OI = 0 \) indicates no occlusion, and \( OI = 1 \) indicates complete occlusion.

Large support threshold values were observed for less occluded leaves \( (OI < 0.75) \), which had appeared in many images, because they had achieved both high precision and recall values by filtering inaccurate 3D curve fragments. Highly occluded leaves \( (OI > 0.75) \) tended to have lower optimal support thresholds at increased \( OI \) values, with the optimal values exhibiting large variations (Fig. 8a), which were attributed to differences in leaf areas, with larger leaves showing steeper trends. Furthermore, the optimal support thresholds decreased when the degree of camera positional error increased (i.e., low positional accuracies prevented precise
filtering) (Fig. 8b) and increased slightly when there were more cameras (Fig. 8c). These trends were clearly observed in leaves with low to intermediate levels of occlusions (0.75–0.80) (Additional file 1: Fig. S4).

Herein, we propose a guideline for determining the support threshold based on simulated data. We conducted Bayesian ridge regression on the optimal support thresholds against the OI based on the simulated data of 128 images with no camera positional noise (Fig. 8d). Moreover, we have provided the following qualitative guidelines: the slope of the linear regression models should be made a downward revision for large leaves (i.e., the trend becomes steeper when the leaf area becomes larger) (Fig. 8a); the camera positional error should be suppressed under a certain value (Fig. 8b); and the number of images should not be increased unnecessarily because improvements in estimation precision reduce when there are more images (Fig. 8c).

Application of proposed method to actual soybean data

We demonstrated the performance of the proposed method by applying it to individual soybeans (Fukuyutaka) at three growth stages (Fig. 9; Additional files 3–5). The 3D leaf edges were reconstructed using the support threshold proposed in the guidelines (Fig. 8d). At the high support threshold, inaccurate reconstructions were suppressed, but some leaves disappeared (Fig. 9, mean). In contrast, at the low support threshold, there was inaccurate reconstruction and the occurrence of artifacts, but almost all the leaves had been reconstructed (Fig. 9, mean – 0.5 SD). It was more challenging to reconstruct all the leaves at a later growth stage because the number of occlusions caused by the number of leaves had increased. Several failed cases were observed: (1) single leaves were reconstructed as multiple leaves because point cloud segmentation had failed in the leaf correspondence step (Fig. 10a); (2) leaves were not reconstructed because the small leaves had disappeared at the mask generation step (Fig. 10b); and (3) reconstructed leaf edges differed markedly from their original shapes owing to B-spline fitting in the cases where they had not been covered by 3D curve fragments (Fig. 10c).

Applicability of the proposed method to holes on a leaf

Using the proposed method, we attempted to reconstruct holes in leaves, which was challenging to do using point-based 3D reconstruction. To demonstrate this, we applied the proposed method to simulate leaves with one to six holes. The edges and holes were reconstructed well when the number of holes was three or fewer (Fig. 11a-c), and the precision of the reconstruction decreased with an increasing number of holes (Fig. 11d-f). Particularly, for leaves with five and six holes, the holes had been reconstructed perpendicular to them (Fig. 11e, f), because the precision of the 3D curve fragment reconstruction had decreased, and the holes had not been separated in a clustering step, in which DBSCAN is used to separate the leaf edge and holes.

Discussion

The proposed phenotyping approach, which includes the instance segmentation of 2D images and curve-based 3D reconstruction that integrates the information into a 3D space, successfully reconstructed 3D leaf edges from multi-view images of both virtual and actual plants (Figs. 4 and 9). Owing to the inclusion of the leaf correspondence identification step, our approach is applicable not only to a single leaf but also to multiple leaves in the same scene (Fig. 5). The direct 3D reconstruction of leaf edges does not require the removal of artifacts from the background (Fig. 2) and allows the robust estimation of leaf edges in a 1D closed curve in 3D space. The simulation results showed that as long as the camera positional errors were not too large (~1 mm), the precision in estimating the leaf edges could be maintained (Fig. 6), even when the number of cameras had been reduced or the degree of occlusion had been changed. Moreover, the proposed method paves the way for solving the problem of point-based 3D reconstruction methods such as SfM/MVS in struggling to distinguish real holes from artifacts (e.g., [36]). The proposed method correctly performs 3D reconstruction only for the holes in leaves recognized in 2D images instead of incorrectly recognizing these holes as the “negative” of the point cloud data. In our simulation, the holes were reconstructed well when the number of holes was less than four. Although the estimation was poor when the number of holes was greater than four, the results could be improved by recognizing each hole as an individual instance in the instance segmentation step, similar to the approach in leaf correspondence identification (Additional file 1: Fig. S2).

To improve the accuracy of 3D edge reconstruction, the following points should be considered: (1) tuning the hyperparameters, (2) improving the camera parameter estimation, and (3) improving the instance segmentation model. These points are elaborated as follows. (1) The hyperparameters used in the proposed method were tuned. We provided the guidelines for setting the support
Regarding SfM, it is a photogrammetric method for simultaneously estimating the camera parameters and the depth of the scene. The SfM technique was utilized to obtain the projection matrix for each camera and the sparse point cloud from a multi-view image.

Conclusions

We have proposed a method for the 3D reconstruction of leaf edges using DNN-based instance segmentation and a curve-based approach from multi-view images to address existing problems in point-based methods. The proposed method reconstructed the 3D leaf edges for both simulated and actual data even in the presence of multiple leaves in the same scene. The accuracy of leaf edge reconstruction was evaluated using simulated virtual plants under different conditions as the ground truth. We inferred that smaller leaves were more challenging to reconstruct than larger ones and that the camera positional error and image numbers did not affect the results that much since they had satisfied certain criteria. Based on the simulation, we have proposed qualitative and quantitative guidelines for setting several parameters used in the proposed method. The proposed method is applicable not only to leaf edges, but also to holes in leaves. We believe that by integrating hierarchical morphological properties, which are difficult to quantitatively evaluate, the proposed method will be a valuable tool for advancing our understanding of plant physiology, functional ecology, and cultivation management.

Methods

Structure from Motion

The SfM technique was utilized to obtain the projection matrix for each camera and the sparse point cloud from a multi-view image. Regarding SfM, it is a photogrammetric method for simultaneously estimating the camera parameters and the depth of...
corresponding points (i.e., sparse 3D point clouds) from multi-view images. In this study, we used Metashape (Agisoft, St. Petersburg, Russia), which is commercial photogrammetry software that includes SFM. The projection matrices, including the optical center, focal length, orientation, and position of the cameras, were exported as Extensible Markup Language (XML) files. Markers were placed on the image to optimize image placement and thereby make it easier to obtain the corresponding points.

Instance segmentation of leaves using Mask R-CNN

To extract the 2D edges of leaves, mask images of multi-view images were obtained using Mask R-CNN, a deep neural network model for segmentation [42]. We used Detectoron2 [56], a library for detection and segmentation tasks, to utilize the Mask R-CNN model with the backbone ImageNet and the model weights pretrained on the COCO dataset. The model was trained on a training dataset that comprised 80% of the dataset consisting of multi-view images of three individual plants, and the remaining 20% of images were used for validation (validation dataset). The model performance was evaluated using a test dataset consisting of another individual plant. To calculate the accuracy of instance segmentation using the Mask R-CNN, we performed four-fold cross-validation, which assigned a different individual as the test dataset.

Leaf edges were extracted from the predicted mask images using OpenCV library [31]. The extracted edges were divided according to the condition [37] under which the contours had random lengths. The 2D edges were divided into fragments with an overlap $\tau_{\text{overlap}}$ and with a certain length $l_{\text{fragment}}$. In this study, we used $\tau_{\text{overlap}} = 15$ and $40 < l_{\text{fragment}} < 100$ for the simulation data and $\tau_{\text{overlap}} = 30$ and $80 < l_{\text{fragment}} < 200$ for the real data.

Leaf correspondence identification across multi-view images

To individually process and reconstruct the leaves, we determined the correspondence of the leaves between the images. First, the point cloud was clustered, and the points in the backside was removed and projected onto the mask image. Then, the point cloud was associated with the mask on which most of the points had been located. If this was performed for all the mask images, the correspondence between the leaves of the images could be obtained via a point cloud. In this study, density-based spatial clustering of applications with noise (DBSCAN) [57], a density-dependent clustering method, was used for clustering on simulated data; and region-growing segmentation implemented in the Point Cloud Library [58], where the angle between normals determines whether a region is grown or not was used on real data. Hidden point removal [39], where the angle between the normals determines whether a region is grown, was used on the real data. Hidden point removal [39], which determines the visible points in a point cloud from a given viewpoint using a sphere and a spherical inversion operator, was used for background removal. Figure 6a shows an overview of the method, and Fig. 6b shows the practical application of the method.

Curve-based 3D reconstruction of leaf edges

A 3D curve sketch [37], which is a curve-based 3D reconstruction, was obtained using the projection matrix obtained by SfM, the multi-view mask image obtained by Mask R-CNN, and the correspondence of leaves among the images. All the subsequent processes were applied to each leaf. Obtaining a 3D curve sketch involves the following steps: (1) camera pair definition; (2) pair hypothesis generation; and (3) 3D curve fragment reconstruction and filtering by reprojection.

(1) Camera pair definition

To perform curve-based 3D reconstruction, camera pairs were defined based on the relative positions of the cameras in the scene. Angle $b_{ij}$, which is the angle between cameras $i$ and $j$ from the average positions of all the cameras ($\bar{b}$), was calculated for all the camera combinations. The camera pairs were defined as the combinations that satisfied. Since the cameras had been assumed to be equally spaced to cover the plants, $\bar{b}$ corresponded to the baseline in [37]. In this study, we used angles of 30, 40, and 60° on the simulated data of 128, 64, and 32 multi-view images, respectively. For the real data, $b_{\text{max}}$ was set to 30°, regardless of the number of images.

(2) Pair hypothesis generation

Let be the $p$-th 2D curve fragment in the $h$-th image. Pair hypothesis, a potentially corresponding pair of 2D curve fragments, is defined as a pair of 2D curve fragments. In epipolar geometry, a fundamental matrix $F_{ij}$ computed from the projection matrices
corresponding to images (and) maps a point in the i-th image to a line in the j-th image. The line mapped by the fundamental matrix is called the epipolar line (or epiline), and any existing corresponding points along the line are found. By extending this concept to a 2D curve fragment, \( F_{ij} \) maps a 2D curve fragment in the i-th image to a band (a set of epipolar lines) in the j-th image. Pair hypotheses were generated based on the 2D curve fragments overlapping the bands (Fig. 3a). For a robust reconstruction, 2D curved fragments tangential to the epipolar line were excluded from the process (see [37] for details). The number of pairs of hypotheses per band was set to a maximum of only 10 to account for the limited computational resources.

(3) 3D curve fragment reconstruction and filtered by reprojection

Then, 3D curve fragments (\( \mathbf{f}_i \)), which correspond to the pair hypotheses, were reconstructed using projection matrixes in 3D Euclidean space. Each reconstructed 3D curve fragment was reprojected onto multi-view images, excluding the i- and j-th images, to evaluate how closely the reconstructed curve fragments generated the true projection. The reconstructed curve fragments were supported by reprojections if the reprojected curve fragments had been located close to the edges of the target object (i.e., leaf) on the image; i.e., a reprojected curve fragment was supported if at least \( \tau_v \) (%) of the curve fragment was located within \( \tau_d \) pixels of the edges in \( \tau_t \) images. Only curves supported with a sufficient number of images (i.e., greater than the support threshold \( \tau_t \)) were reconstructed. A \( \tau_v \) of 80% was used for all cases, and \( \tau_d \) was 11 and 39 pixels for the simulated leaves and actual soybean specimens, respectively.

B-spline closed curve fitting

The 3D curve fragments were integrated into a closed 3D curve by using B-spline fitting. In B-spline fitting, a continuous periodic function is approximated by a piecewise polynomial function, which is a linear combination of the order \( j \) B-spline basis over \( i \)-th interval \( b_{i,j}(l) \) as follows:

\[
f(l) = \mathbf{wb}(l) = ( w_1 \ldots w_{n-1} ) \begin{pmatrix} b_{1,j}(l) \\ \vdots \\ b_{n-k-1,j}(l) \\ b_{n-k,j}(l) + b_{1-k,j}(l) \\ \vdots \\ b_{n-1,j}(l) + b_{0,j}(l) \end{pmatrix}
\]

where \( w_i \) denotes the coefficient of the \( i \)-th B-spline. Based on the coordinate values of the reconstructed 3D curve fragments, the B-spline coefficients were estimated for \( x \), \( y \), and \( z \)-coordinate values using the "curve_fit" function in SciPy [59].

Calculation of optimal support thresholds

For \( \tau_s \), the PR curves in the ground-truth mesh and the reconstructed curve fragment were calculated in the simulation data, which the optimal support threshold is the highest \( \tau_s \) with the highest recall when the precision had exceeded 0.99. The precision is the percentage of ground truths for which the reconstructed curved fragments are within 30 mm, and the recall is the percentage of curved fragments for which the ground truths are within 30 mm. We also excluded points supported by more than \( \tau_p \) on a well-supported curve based on 3D drawings [60], which are modified versions of 3D curve sketches.

The simulation data were comprehensively tested for different precision and recall values with respect to the support threshold, which is defined as the number of images whose ratio to the total number of images (128, 64, and 32 images) falls into 24 equal parts (i.e., between 1 and 0.125). The precision and recall values were recorded; if the precision did not reach one, the minimum value was used as the optimal support threshold.

Virtual plant models
Virtual plant models (single, multiple, and single leaves with holes) were created using Blender (Blender Foundation, Amsterdam, Netherlands). Multi-view images of the virtual plants were generated using Unity (Unity Technologies, San Francisco, CA, US) from 128, 64, and 32 cylindrically arranged views. Three individuals were created based on the multiple-leaf model; each leaf was translated randomly—horizontally from −33.33 to 33.33% and vertically from −14.28–14.29% of the bounding box dimensions—and rotated randomly from −10 to 10°.

Multi-view images of soybean plants

Multi-view images were obtained from four individual soybeans (*Glycine max*), including four cultivars (Enrei, Zairai 51 – 2, Aoakimame, and Saga zairai), to train the Mask R-CNN model and evaluate its performance. These individuals were captured at different growth stages: 1) Enrei: 34 days after sowing (DAS); 2) Zairai 51 – 2: 56 DAS, 3) Aoakimame: 24 DAS; and 4) Saga zairai: 48 DAS. To demonstrate the applicability of the proposed method, multi-view images of another cultivar, Fukuyutaka, at different growth stages of 21, 28, and 42 DAS, were obtained. Each set of multi-view images included 264 images, and approximately 130 images were subsampled. We used a simple fixed photogrammetry system consisting of digital cameras (EOS Kiss X7; Canon, Tokyo, Japan), a turntable (MT320RL40; ComXim, Shenzhen, China), and a camera control application (CaptureGRID4; Kuvacode, Kerava, Finland) (Additional file 1, Fig. S5) to obtain multi-view images.

**Abbreviations**

3D three-dimensional

2D two-dimensional

SfM Structure from Motion

MVS Multi-View Stereo

LiDAR Light Detection And Ranging

SFD standardized Fréchet distance

AP average precision

API AP large

APm AP middle

APs AP small

OI occlusion index

PR Precision-Recall

SD standard deviation

SAM Segment Anything Model

DAS days after sowing

**Declarations**

**Ethics approval and consent to participate**

Not applicable.
Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available in the Zenodo repository, [DOI] and GitHub repository [URL].

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers 20H01381, 21K14947, 22H04727, (to K.N.), Japan Science and Technology Agency (JST) PRESTO Grant Number JPMJPR1605 (to K.N.), JST MIRAI Grant Number JPMJMI20G6 (to K.N.); Moonshot R&D Grant Number JPMJMS2021 (to K.N.), and Bio-oriented technology Research Advancement Institution (BRAIN) Moonshot R&D Grant Number JPJ009237 (to K.N.).

Authors’ contributions

KN conceived and designed this study. HM and KN performed the implementation and analyzed the results. HM and KN were major contributors to writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Horiguchi, R., Inbe, N., and Suzuki, M. for their assistance in making multi-view image datasets.

References


Table 1 is available in the Supplementary Files section.

Figures
Problems in point-based 3D reconstruction. (a) Difficulty in distinguishing the leaf edge from the point cloud data. One of the multi-view images of the leaf (left) and reconstructed point cloud data (right). (b) Holes on the leaf as artifacts in the point-based reconstruction. One of the multi-view images of the leaf without holes (left) and reconstructed point cloud data with missing parts on the leaf (right). (c) Monstera adansonii has holes on the leaves in the growth (left). Example of damaged leaf (Cerasus × yedoensis ‘Somei-yoshino’).
Figure 2

Overview of the proposed method for 3D leaf edge reconstruction. The method reconstructs 3D leaf edges from multi-view images. (a) Each leaf in each image is segmented using Mask R-CNN. (b) Each 2D leaf edge is detected from the segmented leaves. (c) Camera positions and orientations are estimated based on SfM. Simultaneously, sparse point cloud data and projection matrix are obtained for the leaf correspondence step, in which (d) the leaves in the multi-view images are identified. (e) The curve fragments are reconstructed in 3D space using the 3D curve sketch, which integrates the 2D leaf edges, projection matrix, and leaf correspondence. (f) The 3D leaf edges are obtained after fitting closed B-spline curves on each set of 3D curve fragments corresponding to a single leaf.
Figure 3

**Curve-based 3D reconstruction of a leaf edge.** (a) Pair hypotheses are generated in a camera pair by searching for intersecting curve fragments in the 2D images along a band of epipolar lines (blue band). (b) The 3D curve fragments are reconstructed and reprojected onto other images to evaluate how closely the reconstructed curve resembles the true projection. The pair hypothesis is supported if the reprojected 2D curve fragment sufficiently close to the 2D leaf edges (within gray dashed curves). (c) Only the 3D curved fragments supported by a sufficient number of images are reconstructed.
Figure 4

Examples of 3D leaf edge reconstruction on simulated leaves. Reconstructed 3D edges of a single leaf (upper, green) and multiple leaves (lower, gray) using the proposed method. Each reconstructed 3D leaf edge is indicated by a different color. Original meshes (left), reconstructed 3D edges (middle), and overlaid ones (right) are shown.
Figure 5

**Leaf correspondence identification.** (a) Correspondence of leaves between images are identified by projecting the clustered point cloud onto each image. (b) An example of a set of point cloud data clustered into each leaf with the background removal from a particular viewpoint (left) and mask image of the corresponding view (right). (c) Heatmap of the count data of projected point cloud data on a mask image. Peaks indicating the correspondence between clusters and instances in a mask image.
Simulations for evaluating the accuracy of 3D reconstruction. (a) Three levels of occlusions are assumed: no pillars (left), thin pillars (middle), and thick pillars (right). (b) Box plots of SFD against leaf area. Each row and each column correspond to the camera positional noise (top: 0 mm, middle: 1 mm, bottom: 3 mm), and the number of images (left: 32 images, middle 64 images, right: 128 images), respectively. Each color indicates a different level of occlusion (blue: no pillars, orange: thin pillars, and green: thick pillars).
Figure 7

**Mask image generation using Mask R-CNN.** (a) Line plot representing the losses of Mask R-CNN. The validation loss became constant after ca. 7000 epochs. (b) Example of predicted masks of leaves.
Figure 8

Optimal support thresholds proposed based on the simulated leaf data. (a) Scatter diagram of optimal support thresholds against the occlusion index. Each point corresponds to the optimal support threshold that achieves the largest recall when the precision is greater than 0.99. The colors of the markers indicate the leaf area. Box plots of optimal support thresholds for (b) camera positional noise and (c) multi-view images.
Figure 9

**Reconstructed 3D leaf edges reconstructions of the an actual soybean plant.** Each row corresponds to a different growth stage (21, 28 and 42 DAS). Examples of a part of the 2D image of the multi-view images are shown in the left column. Results of 3D leaf edge reconstructions with different support thresholds are shown in the right three columns: the mean – 0.5 SD, the mean – 0.25 SD, and the mean of the predictive distribution of the optimal support threshold to the OI of each leaf. Several failed cases are observed (see Fig. 10 for details).
Three typical failed cases of 3D leaf edge reconstructions. (a) Single leaf reconstructed as multiple leaves. Although a single leaf in the 2D image (right) is observed, the point cloud data of the leaf has been segmented into multiple clusters (middle). Then, two edges are reconstructed based on the leaf (right; blue and beige edges). (b) Leaves have not been reconstructed. A pair of cotyledons are observed in the 2D image (left). They have not been reconstructed because Mask R-CNN has failed to predict them (right). (c) Leaf edge that has been reconstructed far from the actual position. When the 3D curve fragments are not covered over the leaf edge (pink lines: 3D curve fragments), the B-spline curve is overfitted to the boundaries (green line: estimated B-spline curve).
Figure 11

Reconstruction of a leaf edge with hole(s). From upper left to lower right, the number of holes increases from one to six; purple lines represent the reconstructed edges, and the green area the ground-truth mesh. Several holes are reconstructed perpendicular to the leaf in cases with 5 and 6 holes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.xlsx
- Additionalfile1.pdf
- Additionalfile2.gif
- Additionalfile3.mp4
- Additionalfile4.mp4
- Additionalfile5.mp4
• Additionalfile6.xlsx