



OPEN ACCESS

EDITED BY

Yongliang Qiao,
University of Adelaide, Australia

REVIEWED BY

Taewon Moon,
Korea Institute of Science and Technology
(KIST), Republic of Korea
Alen Alempijevic,
University of Technology Sydney, Australia
Yong Suk Chung,
Jeju National University, Republic of Korea

*CORRESPONDENCE

Taek Sung Lee
✉ tslee@kist.re.kr

RECEIVED 27 May 2024

ACCEPTED 04 October 2024

PUBLISHED 24 October 2024

CITATION

Choi H-B, Park J-K, Park SH and Lee TS
(2024) NeRF-based 3D reconstruction
pipeline for acquisition and analysis of
tomato crop morphology.
Front. Plant Sci. 15:1439086.
doi: 10.3389/fpls.2024.1439086

COPYRIGHT

© 2024 Choi, Park, Park and Lee. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

NeRF-based 3D reconstruction pipeline for acquisition and analysis of tomato crop morphology

Hong-Beom Choi, Jae-Kun Park, Soo Hyun Park
and Taek Sung Lee*

Smart Farm Research Center, Korea Institute of Science and Technology (KIST), Gangneung,
Gangwon, Republic of Korea

Recent advancements in digital phenotypic analysis have revolutionized the morphological analysis of crops, offering new insights into genetic trait expressions. This manuscript presents a novel 3D phenotyping pipeline utilizing the cutting-edge Neural Radiance Fields (NeRF) technology, aimed at overcoming the limitations of traditional 2D imaging methods. Our approach incorporates automated RGB image acquisition through unmanned greenhouse robots, coupled with NeRF technology for dense Point Cloud generation. This facilitates non-destructive, accurate measurements of crop parameters such as node length, leaf area, and fruit volume. Our results, derived from applying this methodology to tomato crops in greenhouse conditions, demonstrate a high correlation with traditional human growth surveys. The manuscript highlights the system's ability to achieve detailed morphological analysis from limited viewpoint of camera, proving its suitability and practicality for greenhouse environments. The results displayed an R-squared value of 0.973 and a Mean Absolute Percentage Error (MAPE) of 0.089 for inter-node length measurements, while segmented leaf point cloud and reconstructed meshes showed an R-squared value of 0.953 and a MAPE of 0.090 for leaf area measurements. Additionally, segmented tomato fruit analysis yielded an R-squared value of 0.96 and a MAPE of 0.135 for fruit volume measurements. These metrics underscore the precision and reliability of our 3D phenotyping pipeline, making it a highly promising tool for modern agriculture.

KEYWORDS

3D phenotyping, neural radiance fields, automated growth measurement, point cloud, greenhouse application

1 Introduction

Digital phenotypic analysis is increasingly recognized as an essential element in the accurate morphological analysis of crops and is becoming increasingly important in various application areas (Tripodi et al., 2022). It goes beyond simple observations to digitalize and quantify the crop's genetic trait expressions. In addition, phenotypic analysis is complexly

linked to environmental data, promoting an informed decision-making process to optimize cultivation conditions and improve crop yields.

Two-dimensional (2D) imaging is primarily used in computer vision-based phenotyping studies. For example, segmentation algorithms analyze the number of pixels within a segment to calculate the projected area for area analysis (Fonteijn et al., 2021). Methods, such as the convex hull method, have also been employed to analyze the growth state of crops (Du et al., 2021). Moreover, extracted silhouettes can regressively estimate the crop volume (Concha-Meyer et al., 2018). Despite their usefulness, these methods fail to capture the complete complexity of plant morphology. When reducing 3D structures to 2D representations, significant data that aid the comprehensive understanding of plant health and development, such as leaf curvature, area, and overall plant volume, can be lost.

Several methods using RGB-D cameras and key-point detection (Vit et al., 2019) have been proposed to obtain 3D information about crop morphology. For example, structure from motion-based methods extract phenotyping elements of greenhouse crops from RGB photos captured from multiple angles (Li et al., 2020; Wang et al., 2022). Alternatively, laser scanners can obtain more precise 3D plant models (Schunck et al., 2021). Additionally, obtaining the 3D form of crops is essential for future agriculture, as it enables phenotyping and advanced applications, such as light interception analysis (Kang et al., 2019). However, 3D phenotyping in greenhouse environments poses several challenges. First, even with the light-diffusing film in greenhouses, scattered sunlight can still create substantial noise during measurements using active 3D imaging approaches (such as consumer-level depth cameras like Intel RealSense L515, D435, or laser scanners) (Neupane et al., 2021; Maeda et al., 2022; Harandi et al., 2023). Second, focusing on high-interest positions in greenhouse crops, such as tomatoes, results in low productivity for large and heavy measurement equipment. Third, narrow spacing, typical of greenhouses, makes obtaining sensor data from various angles difficult for 3D model acquisition.

The emerging neural radiance field (NeRF) technology (Gao et al., 2022) offers a new direction for 3D phenotyping. NeRF uses a fully connected neural network to model volumetric scene features and render images from various viewpoints, capturing the 3D structure of a scene. NeRF is robust and can represent complex morphological structures with fewer and more sparsely distributed input images, making it suitable for 3D phenotyping. The time incurred in training NeRFs, which was previously tens of hours, has significantly improved to just minutes with the advent of Instant-NGP, applying hash-encoding-based positional encoding (Müller et al., 2022). Moreover, the user-friendly Nerfstudio framework has made the application and training of NeRF more accessible (Tancik et al., 2023). In agriculture, many applications are underway, including applying semantic segmentation techniques to assist robots' scene understanding in greenhouse (Smitt et al., 2024) or analyzing crops with complex structures (Saeed et al., 2023).

The proposed pipeline encompasses automated RGB image acquisition through a specialized greenhouse robot platform with

a 6-degrees of freedom (6-DoF) robot arm. It also includes acquiring dense point-cloud data utilizing NeRF technology, followed by extracting detailed morphological information from the data. A key aspect of this pipeline is adopting a forward-facing capture technique by operating from a fixed position with a limited field of view, which means that the crops are not captured from a full 360-degree angles at ground level. This limitation aligns more realistically with the practical constraints of greenhouse environments. Despite this limitation, the approach provides the noninvasive measurement accuracy of critical crop parameters, such as length, leaf area, and fruit volume. The application of this method was demonstrated through nondestructive measurements of tomato crops in conditions mirroring actual greenhouse environments. The morphological data obtained were then compared with that acquired through traditional human growth surveys, allowing for a thorough evaluation of the measurement accuracy. The main contributions of this study are as follows:

- A system and data processing pipeline were presented to obtain 3D crop models in greenhouse environments based on images automatically collected by unmanned robots.
- The proposed pipeline demonstrates obtaining decent point-cloud data of crop images from limited viewpoints, showcasing a realistic method for greenhouse environments.
- The proposed 3D plant model was used to measure the following: 1D information, such as stem thickness, node length, and flower position height; 2D information, such as leaf area; 3D information, such as fruit volume. These measurements were compared with actual measurements to demonstrate the suitability of the proposed pipeline for use in growth surveys.

2 Materials and methods

2.1 Pipeline overview

A 3D phenotyping pipeline was presented; 3D point clouds based on RGB images obtained from multiple viewpoints were reconstructed using a robot for nondestructive constraint-overcoming measurements in greenhouse environments. The proposed pipeline comprises seven elements, as shown in Figure 1: (A) Acquiring images using a 6-DoF robot from various viewpoints. (B) Acquiring camera poses from images and calibrating these poses to a meter scale. (C) Training NeRF based on the acquired images and camera poses. (D) Extracting and segmenting the point cloud based on the color and depth rendering results from the NeRF. (E) Skeletonizing to identify connections between plant organs and to extract length information. (F) Reconstructing the surface on the segmented leaf part of the point cloud and calculating the area from the obtained surface. (G) Fitting an ellipsoid to the segmented fruit part of the point cloud and estimating the fruit volume by calculating the ellipsoid volume.

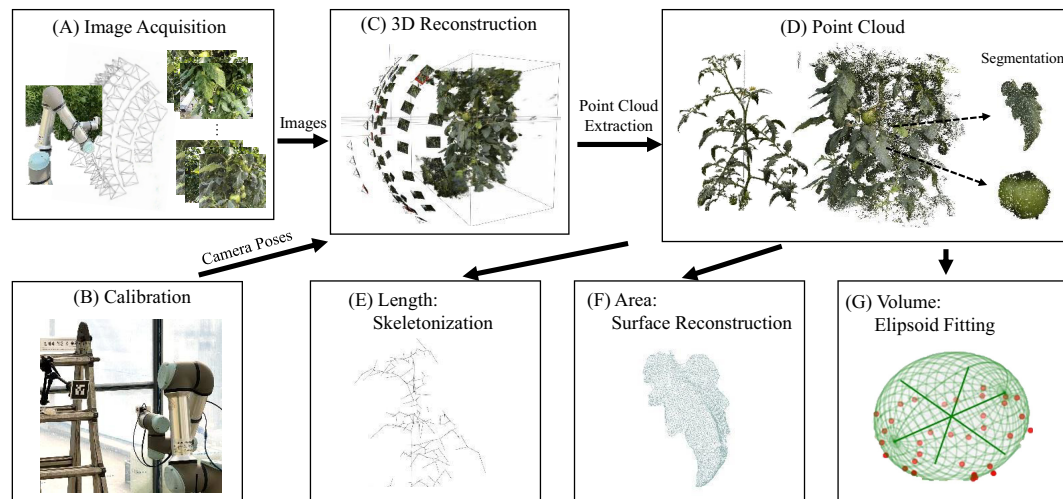


FIGURE 1

3D phenotyping pipeline for tomato crop analysis. The steps are as follows: (A) Image acquisition using a 6-DoF robot, (B) Camera calibration, (C) NeRF training for 3D reconstruction, (D) Point cloud extraction and segmentation, (E) Skeletonization for inter-node length measurement, (F) Surface reconstruction for leaf area measurement, and (G) Ellipsoid fitting for fruit volume estimation.

2.2 Image acquisition

The robot illustrated in Figure 2A was developed in a previous study (Cho et al., 2023) and was employed to facilitate image acquisition in the greenhouse. The base comprises a smart-farm robot platform that controls mobility and provides power. A 6DoF manipulator, UR-5e, with a maximum reach of 850 mm, is mounted on top of this platform. A machine-vision camera, an IDS U3-36L0XC, is attached to the end effector of this robot arm and designed for photographic capture. This camera has a 4200 x 3120-pixel resolution and a frame rate of 20 frames per second. It is

connected to a mini-PC via USB to control the image-capturing process. This mini-PC is connected to the robot arm through a LAN and is equipped with integrated software to control the image-capturing process and robot arm.

In addition, the robot arm-based image acquisition system includes a lift unit for transporting along the Z-axis. This allows maneuvering the robot arm to the desired crop and then vertically to the area of interest using the lift. In the greenhouse environment where the validation was conducted, the average distance between crops was 40 cm, and the distance between lanes was 150 cm. As shown in Figure 2B, the robot was deployed in the field to capture

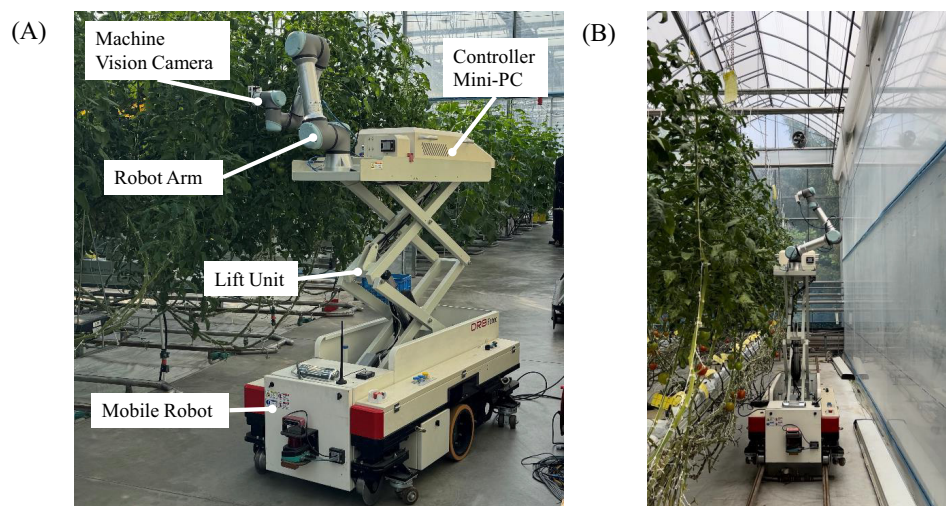


FIGURE 2

(A) Configuration of the greenhouse unmanned robot platform. (B) Example of a robot taking measurements in a greenhouse.

images. The image acquisition process, depicted in [Figure 1A](#), involves capturing images from 64 different poses. These poses are arranged on the surface of a virtual sphere with a radius of 60 cm (the average distance between the crop and the robot arm), centered on the target area of interest. The dome is formed by the portion of the sphere that falls within the robot arm's reach, creating a set of poses that cover the necessary angles.

2.3 NeRF-based 3D reconstruction

NeRF presents a novel approach to 3D scene reconstruction by synthesizing photorealistic images using deep learning. A fully connected neural network is used to model volumetric scene features, rendering complex 3D scenes from 2D image sets. NeRF can interpolate and extrapolate new views from sparse input data, creating highly detailed and coherent 3D reconstructions. The underlying mechanism involves learning the color and density distribution of light in a scene as a function of the position and viewing direction. In NeRF, each pixel is represented as a ray. For each ray, its position information (x, y, z), representing the 3D coordinates in space, and direction (θ, ϕ), representing the viewing angles, are fed into a multilayer perceptron (MLP). The MLP outputs the RGB value and transparency (σ) for that ray. This process essentially captures the light and color information passing through the scene and the density distribution along the rays, providing the basis for reconstructing a 3D model from 2D images. In other words, the NeRF model can receive a 5-dimensional vector, including position and viewing direction, as input and provide RGB values and depth images as output.

Preprocessing to acquire the pose information of the images during the image-acquisition phase is essential for the NeRF to learn and reconstruct a scene. The pose information (x, y, z, θ , and ϕ) is obtained using structure from motion software called COLMAP ([Schonberger and Frahm, 2016](#)). In our pipeline, the UR-5e robot arm, which supports pose repeatability within 0.03 mm, requires running the COLMAP process only once for each set of pre-defined robot arm's poses. The calibration process is shown in [Figure 1B](#).

A marker with known physical measurements were used during calibration. Specifically, a 30 cm by 30 cm printed marker was placed 75 cm away from the robot arm's base to replicate crop measurement conditions. By capturing a single scene with this setup, we were able to obtain the COLMAP results, which provided the marker's coordinates in the reconstructed scene. These coordinates were then matched with the actual known dimensions of the marker, allowing us to determine the scale factor that converts the displacement output (x, y, z) from COLMAP into a metric scale, but also enabled us to reuse the calibrated camera poses in subsequent image captures. As a result, there is no need to recalculate the poses using COLMAP each time, simplifying the NeRF input process.

The **Nerfacto** model within the NerfStudio framework ([Tancik et al., 2023](#)), chosen for its combination of various NeRF-related research advantages, aligns well with the proposed 3D phenotyping pipeline. Despite the excellent pose repeatability of the robot arm,

Nerfacto's camera pose refinement ([Wang et al., 2021](#)) capability is crucial for minimizing potential noise in the results. In addition, hash encoding ([Müller et al., 2022](#)) significantly enhances learning speed, which enhances the overall efficiency of the pipeline. The proposal sampler ([Barron et al., 2022](#)) in Nerfacto, which focuses sample locations on the regions that contribute the most to the final rendering, particularly the first surface intersection, is essential for capturing complex crop details. This focused sampling approach is integral to accurately depicting the intricate morphological traits for detailed phenotypic analysis.

During training, we employed the Nerfacto model in Nerfstudio version 0.3.4, utilizing the default training parameters. However, because we were solely focused on point-cloud acquisition, we did not partition the validation set, and instead, adjusted the number of iterations to 20,000. The training was conducted on a workstation (AMD Ryzen™ Threadripper™ PRO 5975WX, 256GB RAM, NVIDIA RTX4090), and completed in approximately 5 minutes. After training, NeRF's RGB render output and depth map output can be mapped for all camera poses included in the training set and sampled as a point cloud. This process utilized the implementation built into the NeRFStudio framework.

2.4 Phenotypic trait extraction

In this study, we extract phenotypic traits from point clouds generated by an earlier pipeline, focusing on inter-node length, leaf area, and fruit volume. We applied Laplacian-based contraction (LBC) ([Cao et al., 2010](#)) to the point cloud to extract length information, leading to skeletonization. Skeletonization reduces the point cloud to a more manageable representation and emphasizes the structural aspects of the plants. Because the skeleton resulting from the LBC is a collection of discontinuous points, we applied the minimum spanning tree (MST) algorithm ([Meyer et al., 2023](#)) to create a more coherent structure. The MST algorithm transforms the disconnected points into a graph-like structure, effectively representing the plant's physical structure. Thus, the nodes in the skeleton can be aligned with the actual nodes of the crop stem, accurately representing the plant morphology. The final topology graph, extracted from the point cloud, has nodes whose coordinate system corresponds to the original point cloud. Consequently, the Euclidean distance between two points of interest in this graph represents the distance between the crop nodes.

Extracting leaf area and fruit volume measurements requires preprocessing involving point-cloud segmentation. In this pipeline, we manually carried out this segmentation using **CloudCompare** ([2023](#)), as illustrated in [Figure 3](#). Manual segmentation in CloudCompare allows the precise separation of different components of the point cloud, specifically distinguishing leaves, and fruits from other parts of the plant.

To calculate the leaf area, we first performed surface reconstruction on the segmented point cloud of the leaves. The total area was calculated as the sum of the triangular areas forming the mesh obtained from this reconstruction. However, accurately reconstructing a typically thin leaf structure requires noise removal

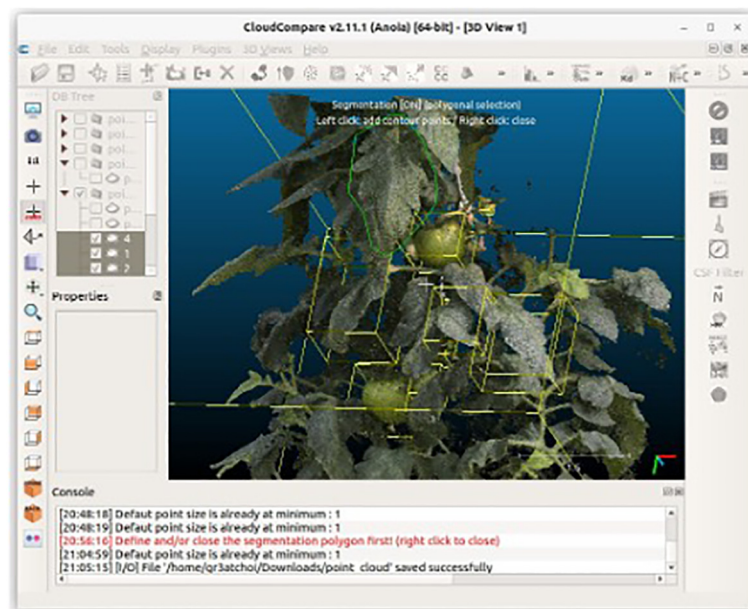


FIGURE 3

Manual segmentation of leaves and fruits using CloudCompare. Segmentation is performed by drawing a polygon (green lines) using the clipping tool. Afterward, a point cloud segmented with a yellow cuboid is displayed.

near the leaf surface. Our pipeline incorporates a moving least squares (MLS) technique to address this (Alexa et al., 2001). MLS effectively converges points near the leaf surface while preserving the natural curvature and shape of the leaves (Boukhana et al., 2022). Next, the ball pivoting algorithm (BPA) was employed to generate the final mesh of the leaf. BPA works by rolling a ball of a specified radius over points to create a mesh, adeptly bridging gaps between points while maintaining the integrity of the leaf's shape.

Finally, we employed ellipsoid fitting to estimate the volume of the segmented tomato fruits. Our pipeline uses input images captured from limited angles rather than from a full 360-degree view, which inevitably limits the measurement of the rear part of the tomato. However, despite this limitation, the tomato volume can be approximated by fitting an ellipsoid to the point cloud representing the measured portion of the tomato.

Ellipsoid fitting in this context is a simple but practical approach for volume estimation when complete data coverage is not feasible (Sari and Gofuku, 2023). By modeling the visible part of the tomato as an ellipsoid, we extrapolate the unmeasured portion, assuming symmetry and typical shape characteristics of tomatoes.

Fitting minimizes the size of the squared distance from the points to the ellipsoid surface, leading to the estimation of semi-axes a , b , and c . The optimization can be represented by minimizing the following function:

$$f(a, b, c) = \sum_{i=1}^M \left(\frac{x_i^2}{a^2} + \frac{y_i^2}{b^2} + \frac{z_i^2}{c^2} - 1 \right)^2 \quad (1)$$

where (x_i, y_i, z_i) are the coordinates of the i -th point in the point cloud, and M is the total number of points in the point cloud. Optimization was performed using the least-squares method.

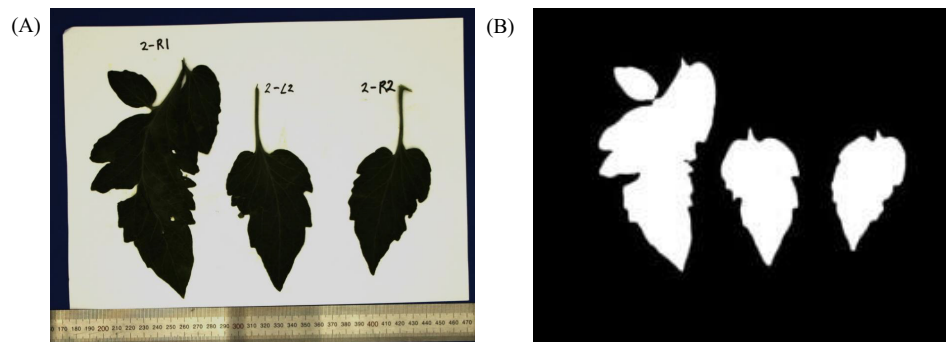


FIGURE 4

Obtained color-image (A) and background removal (B) for leaf area calculation.

Through the optimization, volume V of the ellipsoid fitted to the tomato point cloud can be obtained.

2.5 Ground truth measurement

To evaluate the accuracy of the proposed pipeline, we describe the ground-truth measurement methods conducted alongside image capturing. The results obtained by skilled cultivators using tape measures were used as the ground truth for measuring the

node length. However, considering the node extraction based on skeletonization in our study, measurements were centered on the point where the branches diverged as much as possible.

For the leaf area, leaves were cut, affixed to paper, and photographed in a controlled studio environment, with the camera positioned directly above at a distance of 40 cm, ensuring a perpendicular angle. An example of a photographed leaf is shown in Figure 4A. The ruler is included to facilitate the conversion between pixel units and metric scale. Subsequently, binary processing was applied to these images to create silhouettes of the



leaves, as shown in Figure 4B. The leaf area was then determined by calculating the pixel area of the silhouette in square centimeters (cm^2) using a scale factor obtained from 1 cm pixels on the ruler.

Finally, for volume measurement, we utilized the principle of buoyancy, which calculates the volume of an object based on the weight and force required to submerge it in water (Concha-Meyer et al., 2018). The weight [g] was measured using a scale with a resolution of 0.05 g. A glass beaker filled with water was placed on the scale, and its tare function was used to adjust the reading to zero. The fruit, attached to a wire, was quickly submerged in water and positioned at the center of the beaker. The reading, which reflects the weight of the submerged fruit and the weight when pressed down by the wire, was recorded and represents the volume of the fruit [cm^3].

3 Results

For validation, we measured the growth points of 16 tomato crops at the upper parts and the fruit clusters of 16 tomato crops at the lower parts, resulting in 32 image sets, each comprising 64 multi-view images. From these, we measured a total of 47 inter-node lengths, including 1 inter-node length above and 1-2 inter-node lengths below the topmost flower cluster for each plant. In each of the lower part image sets, we measured 2-3 leaves and 1-2 fruits. These selections were based on factors such as size, shape, and proximity to the robot arm to ensure diversity, resulting in measurements of 37 leaf areas and 20 fruit volumes in total. All measurements were conducted concurrently with ground-truth measurement.

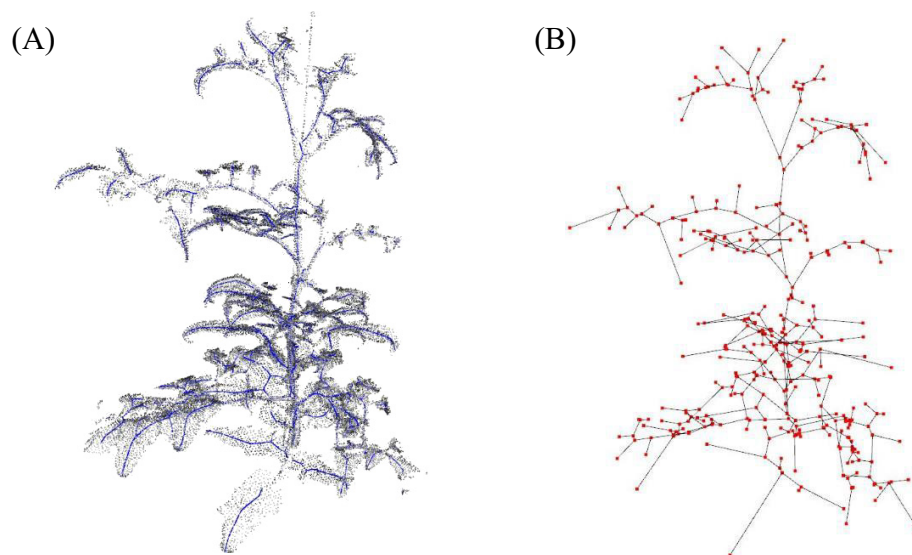


FIGURE 6

(A) Skeletonization process through applying Laplacian-based contraction to point cloud and (B) applying minimum spanning tree algorithm to the skeleton; blue and red dots indicate skeleton and nodes, respectively.

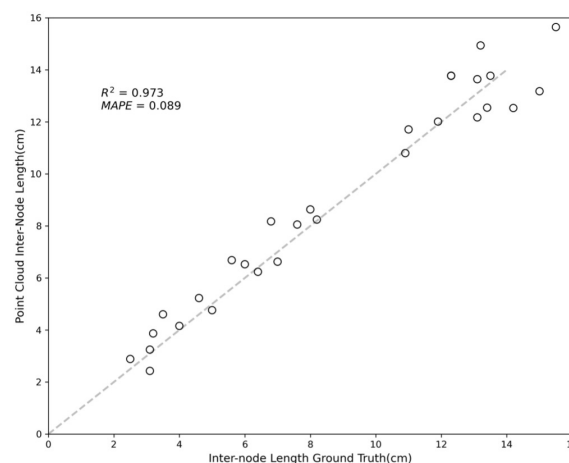


FIGURE 7

Comparison between inter-node length from NeRF and ground truth.

Figure 5 shows the extracted point clouds, illustrating two growth points and two fruit clusters. The front-view representation displays a dense formation of the point cloud, as captured from the angle at which the images were captured. However, the side view, representing angles not captured during imaging, shows reduced performance, especially in regions not directly imaged.

Figure 6 presents the skeletonization results performed to measure the node-to-node lengths. In Figure 6A, the skeleton created through LBC is overlaid on the original point-cloud data as blue dots. Figure 6B shows the application of the MST algorithm to this skeleton; the red dots represent nodes, and the connections between them are depicted.

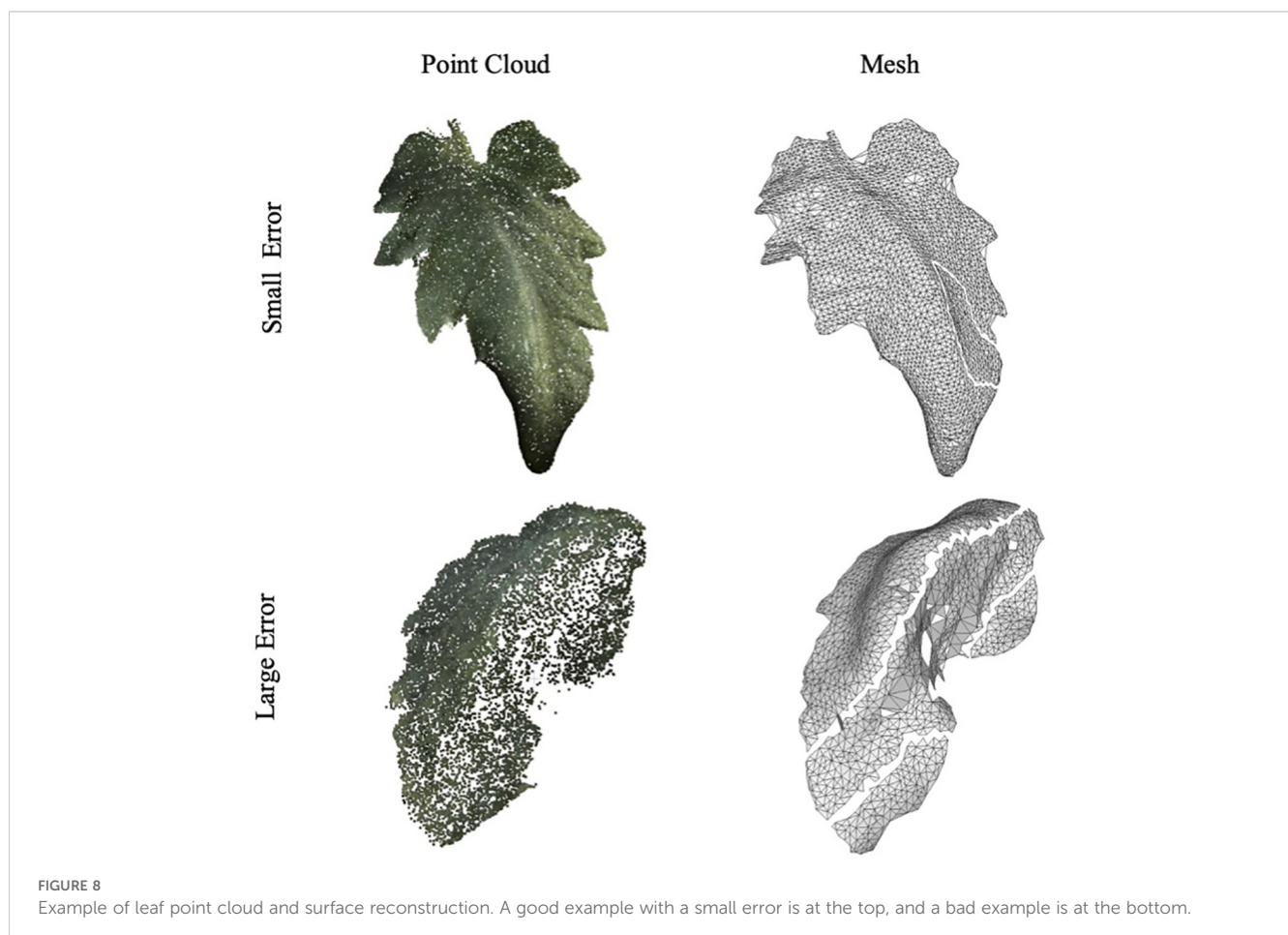
A comparison between the distances measured among the red dots and the node-to-node lengths measured manually is shown in Figure 7. The results showed an R-squared value of 0.973 and a mean absolute percentage error (MAPE) of 0.089, indicating a accuracy in the skeletonization and subsequent measurements. The error sources can be attributed to fundamental differences in the measurement approaches; the point cloud data measure lengths based on the central coordinates of the plant nodes, while the tape measure records lengths over the plant's surface. The discrepancy between the two methods may account for the minor measurement variations. Despite these differences, the close correlation demonstrates the effectiveness of the skeletonization process in accurately capturing the crop's physical dimensions.

Figure 8 showcases examples of the segmented leaf point cloud and the meshes reconstructed using MLS and BPA. The top two samples demonstrate instances with minimal error, serving as representative examples of high accuracy, while the bottom two samples exhibit significant discrepancies, highlighting cases with large errors.

The variance in accuracy between these two sets of samples is attributed to the image-capturing angle. Samples with greater errors include leaves that were curled or rolled up, resulting in one side not being adequately captured. This lack of complete data led to inaccuracies in the reconstruction process. Figure 9 further illustrates this point, with an R-squared value of 0.953 and a MAPE of 0.090.

Figure 10 shows examples of segmented tomato fruits. Similar to the previous examples with leaves, the top two samples represent instances with minimal errors, whereas the bottom two samples show substantial discrepancies. The high errors likely resulted from the fruits being partially obscured by leaves or adjacent fruits, leading to fewer data points being captured and, consequently, errors in the fitting process.

In Figure 11, the results are quantified, showing an R-squared value of 0.96 and a MAPE of 0.135. These values indicate a high degree of accuracy in most cases, with errors primarily arising from occluded or partially hidden portions of the fruits.



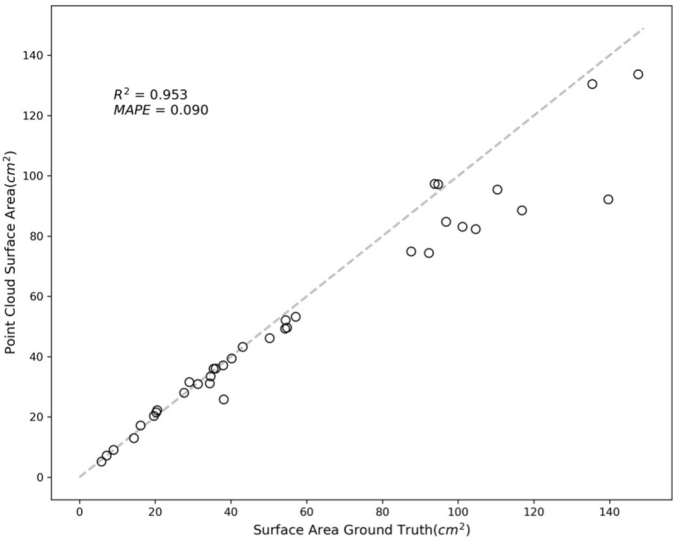


FIGURE 9
Comparison of surface area from NeRF and ground truth.

The issues encountered, predominantly owing to obscured parts of the fruit, suggest a potential avenue for improvement in future studies. Addressing this challenge may involve fitting parametric geometric models to the fruits or implementing appropriate interpolation

methods. Such techniques can help accurately estimate the shape and volume of partially obscured fruits, enhancing the phenotyping precision. This approach is beneficial in complex agricultural environments where occlusion by leaves or other fruits is common.

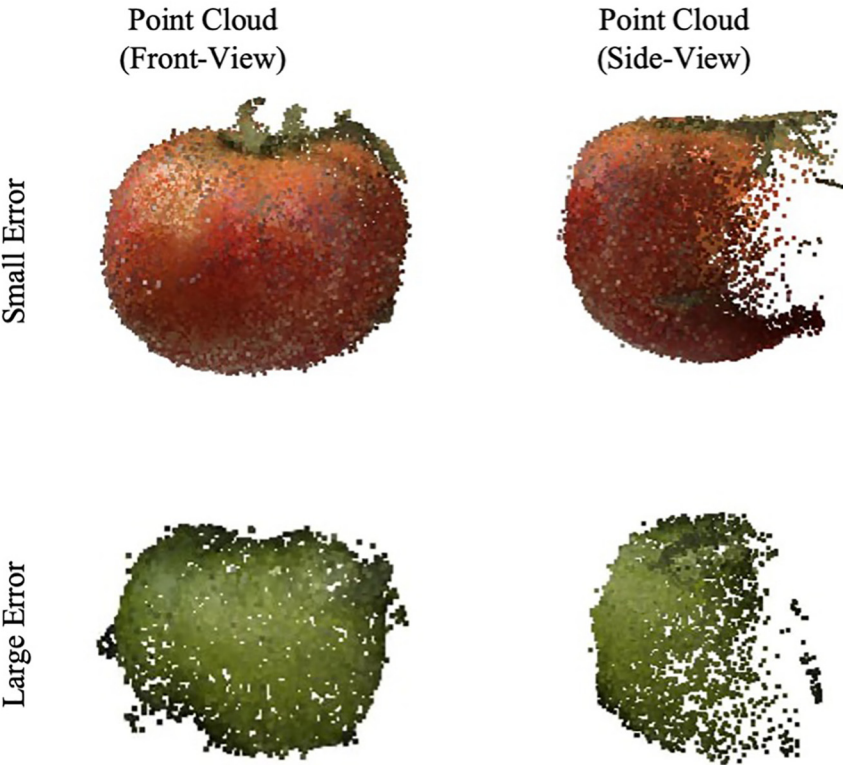


FIGURE 10
Example of tomato fruit's point cloud. A good example with a small error is at the top, and a bad example is at the bottom.

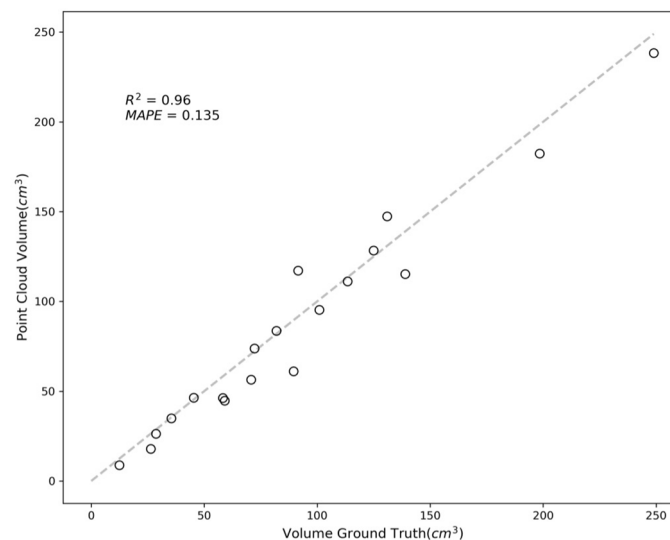


FIGURE 11
Comparison between volume from NeRF and ground truth.

4 Discussion

4.1 Performance

The 6-axis robotic arm mounted on the SmartFarm robotic platform used in this study can capture images at 64 poses in approximately 240 seconds. Due to its general-purpose design, not specifically optimized for capturing images quickly, the image acquisition process is relatively time-consuming. However, the development of dedicated image acquisition hardware could significantly enhance both image acquisition speed. For example, implementing a system with rails capable of Z-axis movement and a camera unit with Pan-Tilt functionality could be considered. This system can perform multi-angle image acquisition in conjunction with the robot's movement on the ground. This approach will not only reduce the time required for multi-angle image acquisition but also lower hardware costs.

4.2 Multi-modal point cloud

If an additional camera, such as an IR or multispectral camera, is installed parallel to the RGB camera when acquiring images, it becomes feasible to implement a multimodal point cloud. In the NeRF point cloud construction process mentioned above, the point cloud is sampled by mapping the RGB image and depth image generated as the output of NeRF. By aligning multimodal images taken from the same angle with the generated RGB image, it is possible to obtain a point cloud where the color from RGB is replaced by multimodal values. Specifically Thermal imaging using IR camera can be used to extract physiological indicators from plants (Pradawet et al., 2023). However, when these thermal data

are integrated into a 3D structure, enabling the point cloud to include both morphological and physiological information, there is potential to develop more sophisticated indicators for analyzing plant stress or disease, which could lead to more accurate and representative plant physiological assessments.

4.3 Limitations and future work

Our current study, while demonstrating the potential of NeRF-based 3D reconstruction for tomato crop phenotyping, has several limitations that need to be addressed in future research. One limitation is that the process of extracting regions of interest from the point cloud or node graph is currently manual. This manual process introduces the potential for human error and limits scalability. To enable high-throughput phenotyping, it will be essential to incorporate additional technologies, such as AI-driven 3D segmentation (Xie et al., 2020), which could automate this process and significantly improve efficiency.

Another limitation lies in the image acquisition process, which is constrained by the robot's fixed position, capturing images only from angles within the reach of the robotic arm. While this method has proven effective for capturing the structure of plant nodes in the upper parts of the crops, it struggles with densely vegetated lower parts where leaves and other foliage can obscure key details. The occlusion effects caused by tightly packed leaves can result in sparse point clouds and reduced accuracy in the final 3D model. The fine details of smaller leaves are particularly prone to being smoothed out or lost during the reconstruction process, further complicating accurate representation. To address these challenges, one approach could involve applying models that can efficiently process more numerous and detailed input images (Wang et al., 2022), thereby capturing finer details and reducing

occlusion issues. Another potential solution is to integrate autonomous navigation technologies such as SLAM (Simultaneous Localization and Mapping). By using SLAM (Campos et al., 2021), the robot could link image sets captured from different locations, such as combining images taken from the opposite lane of the target crop, to provide a more complete view.

Additionally, the current pipeline is specifically designed for tomato crops in a greenhouse environment, with limitations in accurately measuring fruit volumes in the lower parts of the plants. Expanding the applicability of this pipeline will require more robust 3D data acquisition methods, possibly through enhanced image coverage using autonomous mobility or more sophisticated interpolation techniques, to provide comprehensive volumetric data.

Lastly, we encountered challenges related to crop movement during image capture in real greenhouse environments. Even slight movements of the crops during shooting introduced noise into the 3D reconstruction results, which compromised accuracy. To address this vulnerability, applying dynamic NeRF (Pumarola et al., 2021) that add a time axis to the radiance fields could allow the system to capture the geometry of moving objects. If integrated into 3D phenotyping, this approach could enable the system to operate robustly even in open fields where wind and crop movement are factors, offering a promising direction for future research.

5 Conclusion

By employing a state-of-the-art combination of NeRF technology and autonomous robotic systems, we successfully developed a pipeline capable of capturing comprehensive morphological crop data from limited viewpoints. The precision of our method was validated by the R-squared values above 0.953 and MAPE under 0.96 for length, area, and volume measurements, demonstrating its superiority over traditional growth surveys. However, our study identified challenges, such as occlusion and incomplete data capture due to foliage, indicating areas for future enhancement. Potential improvements could involve integrating parametric geometric modeling or sophisticated interpolation methods for more accurate shape and volume estimations of partially visible fruits. Overall, this research proves the viability of advanced 3D phenotyping in real-world greenhouse scenarios and paves the way for future developments in digital agriculture to optimize crop management and yield through precise morphological assessments.

References

- Alexa, M., Behr, J., Cohen-Or, D., Fleishman, S., Levin, D., and Silva, C. T. (2001). "Point set surfaces," in *Proceedings Visualization, 2001 VIS '01*, San Diego, CA, USA. (Washington, DC, United States: IEEE), 21–537. Available at: <http://ieeexplore.ieee.org/document/964489/>.
- Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., and Hedman, P. (2022). "Mip-NeRF 360: unbounded anti-aliased neural radiance fields," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (New Orleans, LA, USA: IEEE), 5460–5469. Available at: <https://ieeexplore.ieee.org/document/9878829/>.
- Boukhana, M., Ravaglia, J., Hétroy-Wheeler, F., and De Solan, B. (2022). Geometric models for plant leaf area estimation from 3D point clouds: A comparative study. *Graph Vis. Comput.* 7, 200057. doi: 10.1016/j.gvc.2022.200057
- Campos, C., Elvira, R., Rodriguez, J. J. G., Montiel, J. M., and D. Tardos, J. (2021). ORB-SLAM3: an accurate open-source library for visual, visual-inertial, and multimap SLAM. *IEEE Trans. Robot.* 37, 1874–1890. doi: 10.1109/TRO.2021.3075644
- Cao, J., Tagliasacchi, A., Olson, M., Zhang, H., and Su, Z. (2010). "Point Cloud Skeletons via Laplacian Based Contraction," in *2010 Shape Modeling International*

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

H-BC: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Data curation. J-KP: Data curation, Software, Writing – original draft, Investigation. SP: Methodology, Writing – review & editing, Investigation. TL: Conceptualization, Project administration, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry and the Korea Smart Farm R&D Foundation (KosFarm) through the Smart Farm Innovation Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs, Ministry of Science and ICT, and Rural Development Administration (421025-04).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer TM declared a shared affiliation with the authors to the handling editor at the time of review.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Conference. (Aix-en-Provence, France: IEEE), 187–197. Available at: <http://ieeexplore.ieee.org/document/5521461/>.

Cho, S., Kim, T., Jung, D. H., Park, S. H., Na, Y., Ihn, Y. S., et al. (2023). Plant growth information measurement based on object detection and image fusion using a smart farm robot. *Comput. Electron Agric.* 207, 107703. doi: 10.1016/j.compag.2023.107703

CloudCompare. (2023). *CloudCompare*. Available online at: <http://www.cloudcompare.org/> (accessed December 23, 2023).

Concha-Meyer, A., Eifert, J., Wang, H., and Sanglay, G. (2018). Volume estimation of strawberries, mushrooms, and tomatoes with a machine vision system. *Int. J. Food Prop.* 21, 1867–1874. doi: 10.1080/10942912.2018.1508156

Du, J., Fan, J., Wang, C., Lu, X., Zhang, Y., Wen, W., et al. (2021). Greenhouse-based vegetable high-throughput phenotyping platform and trait evaluation for large-scale lettuces. *Comput. Electron Agric.* 186, 106193. doi: 10.1016/j.compag.2021.106193

Fontein, H., Afonso, M., Lensink, D., Mooij, M., Faber, N., Vroegop, A., et al. (2021). Automatic phenotyping of tomatoes in production greenhouses using robotics and computer vision: from theory to practice. *Agronomy* 11, 1599. doi: 10.3390/agronomy11081599

Gao, K., Gao, Y., He, H., Lu, D., Xu, L., and Li, J. (2022). NeRF: neural radiance field in 3D vision, A comprehensive review. *arXiv*. Available at: <http://arxiv.org/abs/2210.00379>.

Harandi, N., Vandenbergh, B., Vankerschaver, J., Depuydt, S., and Van Messen, A. (2023). How to make sense of 3D representations for plant phenotyping: a compendium of processing and analysis techniques. *Plant Methods* 19, 60. doi: 10.1186/s13007-023-01031-z

Kang, W. H., Hwang, I., Jung, D. H., Kim, D., Kim, J., Kim, J. H., et al. (2019). Time change in spatial distributions of light interception and photosynthetic rate of paprika estimated by ray-tracing simulation. *Prot Hortic. Plant Fact.* 28, 279–285. doi: 10.12791/KSBEC.2019.28.4.279

Li, D., Shi, G., Kong, W., Wang, S., and Chen, Y. (2020). A leaf segmentation and phenotypic feature extraction framework for multiview stereo plant point clouds. *IEEE J. Sel. Top. Appl. Earth Obs Remote Sens.* 13, 2321–2336. doi: 10.1109/JSTARS.4609443

Maeda, N., Suzuki, H., Kitajima, T., Kuwahara, A., and Yasuno, T. (2022). Measurement of tomato leaf area using depth camera. *J. Signal Process.* 26, 123–126. doi: 10.2299/jsp.26.123

Meyer, L., Gilson, A., Scholz, O., and Stamminger, M. (2023). “CherryPicker: semantic skeletonization and topological reconstruction of cherry trees,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. (Vancouver, BC, Canada: IEEE), 6244–6253. Available at: <https://ieeexplore.ieee.org/document/10208826/>.

Müller, T., Evans, A., Schied, C., and Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.* 102, 1–15. doi: 10.1145/3528223.3530127

Neupane, C., Koirala, A., Wang, Z., and Walsh, K. B. (2021). Evaluation of depth cameras for use in fruit localization and sizing: finding a successor to kinect v2. *Agronomy* 11, 1780. doi: 10.3390/agronomy11091780

Pradawet, C., Khongdee, N., Pansak, W., Spreer, W., Hilger, T., and Cadisch, G. (2023). Thermal imaging for assessment of maize water stress and yield prediction under drought conditions. *J. Agron. Crop Sci.* 209, 56–70. doi: 10.1111/jac.v209.1

Pumarola, A., Corona, E., and Pons-Moll, G. (2021). “Moreno-noguer F. D-neRF: neural radiance fields for dynamic scenes,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Nashville, TN, USA: IEEE), 10313–10322. Available at: <https://ieeexplore.ieee.org/document/9578753/>.

Saeed, F., Sun, J., Ozias-Akins, P., Chu, Y. J., and Li, C. C. (2023). “PeanutNeRF: 3D radiance field for peanuts,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. (Vancouver, BC, Canada: IEEE), 6254–6263. Available at: <https://ieeexplore.ieee.org/document/10209004/>.

Sari, Y. A., and Gofuku, A. (2023). Measuring food volume from RGB-Depth image with point cloud conversion method using geometrical approach and robust ellipsoid fitting algorithm. *J. Food Eng.* 358, 111656. doi: 10.1016/j.jfoodeng.2023.111656

Schönberger, J. L., and Frahm, J. M. (2016). “Structure-from-motion revisited,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Las Vegas, NV, USA: IEEE), 4104–4113. Available at: <http://ieeexplore.ieee.org/document/7780814/>.

Schunck, D., Magistri, F., Rosu, R. A., Cornelißen, A., Chebrolu, N., Paulus, S., et al. (2021). Pheno4D: A spatio-temporal dataset of maize and tomato plant point clouds for phenotyping and advanced plant analysis. *Agudo A editor. PLoS One* 16, e0256340. doi: 10.1371/journal.pone.0256340

Smitt, C., Halstead, M., Zimmer, P., Läbe, T., Guclu, E., Stachniss, C., et al. (2024). PAgneRF: towards fast and efficient end-to-end panoptic 3D representations for agricultural robotics. *IEEE Robot Autom Lett.* 9, 907–914. doi: 10.1109/LRA.2023.3338515

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Kerr, J., et al. (2023). “Nerfstudio: A modular framework for neural radiance field development,” in *ACM SIGGRAPH 2023 Conference Proceedings*. New York, NY: Association for Computing Machinery.

Tripodi, P., Nicastro, N., and Pane, C. (2022). Digital applications and artificial intelligence in agriculture toward next-generation plant phenotyping. *Cammarano D editor. Crop Pasture Sci.* 74, 597–614. doi: 10.1071/CP21387

Vit, A., Shani, G., and Bar-Hillel, A. (2019). “Length phenotyping with interest point detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. New York, NY: IEEE.

Wang, Y., Hu, S., Ren, H., Yang, W., and Zhai, R. (2022). 3DPhenoMVS: A low-cost 3D tomato phenotyping pipeline using 3D reconstruction point cloud based on multiview images. *Agronomy* 12, 1865. doi: 10.3390/agronomy12081865

Wang, C., Wu, X., Guo, Y. C., Zhang, S. H., Tai, Y. W., and Hu, S. M. (2022). “NeRF-SR: high quality neural radiance fields using supersampling,” in *Proceedings of the 30th ACM International Conference on Multimedia*. (Lisboa Portugal: ACM), 6445–6454. doi: 10.1145/3503161.3547808

Wang, Z., Wu, S., Xie, W., Chen, M., and Prisacariu, V. A. (2021). NeRF-: Neural Radiance Fields Without Known Camera Parameters. Available online at: <https://arxiv.org/abs/2102.07064> (accessed December 8, 2023).

Xie, Y., Tian, J., and Zhu, X. X. (2020). Linking points with labels in 3D: A review of point cloud semantic segmentation. *IEEE Geosci Remote Sens Mag.* 8, 38–59. doi: 10.1109/MGRS.6245518