

Effect on Quality of 3D Model of Plant with Change in Number and Resolution of Images Used: An Investigation



Abhipray Paturkar, Gourab Sen Gupta, and Donald Bailey

Abstract 3D reconstruction of plants is a challenging task. It is essential in non-destructive plant growth monitoring systems and important to reconstruct plant features precisely. Two parameters which critically affect quality of the 3D model are change in the number of input images and the resolution of the images. In this paper, the effect on 3D models of these parameters is analysed. This analysis enables an appropriate number of input images and resolution of the images to be determined for a precise 3D model. To validate the results, the plant stem height and number of leaves extracted from reconstructed 3D models were compared with ground truth values. We used descriptive statistics for validation and achieved high correlation between extracted and ground truth values.

Keywords 3D reconstruction · Structure-from-motion · Feature extraction · Feature matching · Plant phenotyping

1 Introduction

3D models provide information about plant status like plant growth or plant disease for agricultural management [1]. 3D plant models can be used to estimate plant features or parameters to avoid the subjective biases associated with human evaluation [2]. Plant features such as stem height, number of leaves, leaf area and so on can be extracted accurately from the 3D model which can be extremely important in the decision making process. These 3D models can characterize plants with complex architecture, supplying important data to plant breeding plans that is inevitably essential for altering features related to plant stress, shape or agricultural management [3]. 3D modeling makes data, such as plant growth monitoring or treatment, available to farmers or growers. In addition, 3D models can be helpful to farmers in yield estimation, disease detection, weed and crop discrimination, and to describe

A. Paturkar (✉) · G. S. Gupta · D. Bailey
School of Food and Advanced Technology, Massey University, Palmerston North, New Zealand
e-mail: A.Paturkar@massey.ac.nz

© Springer Nature Singapore Pte Ltd. 2021
S. N. Merchant et al. (eds.), *Advances in Signal and Data Processing*,
Lecture Notes in Electrical Engineering 703,
https://doi.org/10.1007/978-981-15-8391-9_28

377

plant quality. Even though 3D modeling provides the details of plant structure, it still requires technological advancements for capturing images and extracting plant features.

The majority of modeling techniques are based on 2D images, such as hyperspectral and thermal imaging [4]. Reconstruction in 3D is rapidly getting attention and has recently been the focus of much research. Two main classes of 3D measurement are active and passive techniques. LiDAR [5] and structured light [6] are active sensing methods which use their own source of illumination for sensing. In comparison, passive sensing methods which use radiation in the scene (illumination from the sun) include, stereo vision [7] structure-from-motion [8] and many more [4]. Most passive sensing methods use one or two cameras for sensing which makes them economical compared to active sensing methods.

LiDAR measures the distance from the sensor to each part of the object by measuring the time it takes for a laser pulse to return to the sensor. Paulus et al. [9] presented reliability of LiDAR to obtain an accurate 3D point cloud of a plant but the model gave limited information about surface area because of poor LiDAR resolution. Kaminuma et al. [10] successfully reconstructed a plant model in 3D which represented leaves as a polygonal mesh and used this 3D model for plant feature extraction. Even though LiDAR has performed well in reconstructing plants in 3D, it has several disadvantages, such as, high warm-up time, poor resolution, being very costly and needs numerous captures to handle occlusion [4]. Another method to get depth information is to use structured light (one example is the Kinect sensor). This sensor projects a known pattern on the object and deformation of this pattern allows the vision system to infer surface and depth information of the object [11]. Chéné et al. [12] used a Kinect sensor for 3D plant reconstruction and achieved good results but this struggled to perform in an outdoor environment because of poor contrast of the projected pattern due to bright sunlight. In addition, they did not consider change in any parameters for reconstruction process such as, the number of input images or change in image resolution. Baumberg et al. [13] also presented a 3D reconstruction system for a controlled environment. They used a mesh processing approach for generating 3D models of cotton plants and did not consider change in the reconstruction parameters.

In contrast, passive sensing methods use radiation in the scene, which allows this method to work in outdoor conditions efficiently. Stereo vision has several parameters which the user has to keep in mind, such as distance between two cameras, distance between camera and the plant, and the focal length of the cameras [14]. This can make stereo vision systems difficult to set up and use as the user has to adjust these parameters. Takizawa et al. [15] used a stereo vision system to generate a 3D model of plants and extracted plant shape and height data. There's a still scope for experimentation in this system by considering a few more plant features like number of leaves, leaf area, leaf shape. Ivanov et al. [16] extracted plant structure information using a stereo vision system in outdoor conditions. They captured plant images from different angles and views. They generated a 3D model by considering all the camera views for precise reconstruction.

In structure-from-motion (SfM), set of 2D images are used to create a 3D model of an object. Unlike stereo vision, SfM uses a single camera. A 3D model is generated using SfM by capturing a series of images as the camera is moved around the plant. SfM calculates the distance between camera and plant by itself and it does not require any prior camera calibration, this makes SfM easier to use [4]. Jay et al. [17] presented a system to generate a 3D model of plants in a row and then extract the parameters from the 3D model. The camera was translated along the row which gives limited information about the plant, plant height and leaf area information because of occlusion of leaves. Quan et al. [18, 19] used SfM to reconstruct a tree in 3D. It was a semi-automatic system in which user can select which images are used for the reconstruction. However, the authors did not consider the change in number of input images for the generating 3D model. In this research, we will be using structure-from-motion method for reconstruction because of its easier implementation.

1.1 3D Recovery of Plants

Precise reconstruction of an object from multiple images is an on-going problem. In the past few years advancements to these reconstruction techniques have been made. However, these techniques have been applied to simpler objects e.g. human face, vases, buildings or some round objects. Objects with more complex architecture such as, plants, pose more challenges and issues to precisely reconstruct in 3D. Complex architectures are subject to significant occlusion, where a leaf is not visible from current view, and the parallax effect, where plant appearance differs from view to view, making reconstruction of plants more difficult than convex objects. Reconstruction of plants is difficult because of high self-occlusion, presence of many leaves, shiny surfaces, and texture-loss in some camera views making feature matching more difficult. In addition, plants are very sensitive to changes in the environment, right from small changes like foliage reorganisation to life long growth patterns. Consequently, plants have complex architecture which changes over time which makes it complicated to reconstruct especially by fixed, standard camera phenotyping systems. To reduce this complication and to contribute to the solution of this problem we proposed an easy and cost-effective plant 3D reconstruction system in our previous work [1]. However, we determined that there are some more parameters which need attention to reconstruct a precise 3D model such as the number of input images used for reconstruction and resolution of the images. The speed and quality of a reconstructed 3D model largely depends on the number of input images and selection of the views. It is not necessary that every image will contribute equally to overall quality of a 3D model. It is difficult to select the number of images because with a limited number of images it is difficult generate an accurate plant model because fewer images might generate false points in the 3D point cloud. In contrast, a large number of images will result in processing redundant information which will inevitably increase computation time [20, 21]. To get appropriate number of images

for precise 3D reconstruction, prior information or manual measurements of plants are required to compare to the corresponding measurements from the 3D model.

Clearly, there is a need to consider the number of input images and their resolution for the reconstruction process. These parameters have not yet been discussed in this area of plant phenotyping. In this work, we investigate these two parameters. Our contributions are: (1) To reconstruct plant in 3D using a different number of input images followed by extraction of plant feature, such as stem height and number leaves in 3D. These extracted 3D values are then compared to manual measurements of the plant using descriptive statistics to determine an appropriate number of images for reconstruction. (2) To analyse the effect of image resolution on the quality of the 3D data.

The remainder of the paper is as follows; Sect. 2 discusses the detailed material and methods used in this paper to analyse the effects of the investigation parameters on the quality of the 3D model. The experimental results are discussed in Sect. 3.

2 Materials and Method

For this experiment, we considered a chilli plant (*Capsicum annum L.*) and conducted this experiment on a commercial farm in Palmerston North, New Zealand. The appropriate permissions from the responsible authorities of the commercial farm for conducting this experiment were taken. The chilli plant was selected because of its demand throughout the year and its high value. The stem height of the plant was 15 cm with 11 leaves on it. We aimed at reconstructing a single plant in 3D; hence, the plants were planted in a row having a distance of approximately 90 cm between adjacent plants. Therefore, other plants did not interfere in the 3D model and only one plant is modelled.

2.1 Image Acquisition

The image acquisition process and sample acquired images are shown in Fig. 1. Images are captured while moving the camera around the plant in an approximately circular path. We captured the images in six rounds at different views, angles, distances, and heights. 15 images were acquired in every round [1], with images acquired at approximately 10° – 15° intervals. The 90 images from these six rounds had an overall 90% overlap between images. The distance between camera and plant was variable. The experiment was performed in outdoor conditions which gave varying image quality, such as images with shadow, wind, change in sunlight because of cloud movement. Structure-from-motion calculates the camera intrinsic and extrinsic parameters automatically; hence, different positions of the camera do not need any calibration process.

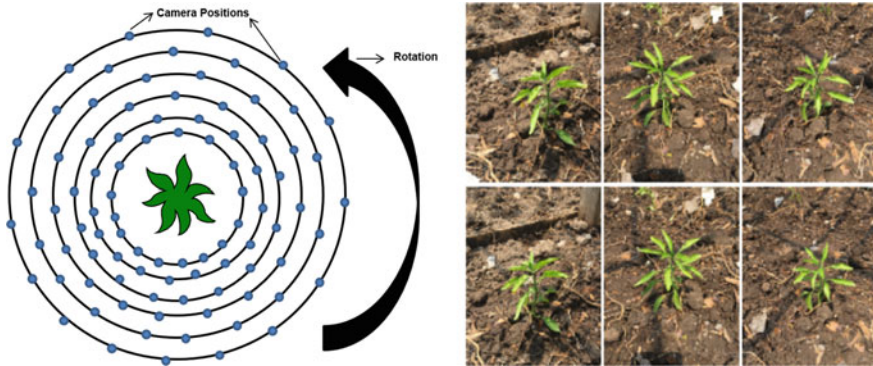


Fig. 1 Image acquisition process and sample acquired images

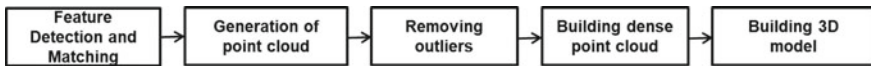


Fig. 2 3D modeling pipeline

2.2 3D Modeling

Once the images, have been acquired, it is necessary to detect common keypoints and then match these keypoints between other camera view images. Figure 2 shows the 3D modeling pipeline. A scale-invariant feature transform (SIFT) [22] was used to detect the keypoints.

1. **Scale-space extrema detection:** The primary step of SIFT searches over all image locations and scales using a difference-of-Gaussian function to detect promising keypoints that are orientation and scale invariant.
2. **Localisation of keypoints:** Keypoints are filtered to remove those with poor stability. Stability is a measure of the sensitivity of keypoints to changes in position and scale.
3. **Orientation assignment to the keypoints:** Orientations are allocated to every keypoint depending on local image gradient directions. These are calculated based on the detected scale to give scale invariance.
4. **Keypoint descriptor:** Local image gradients are calculated at the chosen scale in the area around every keypoint. These gradients are then transformed into a description which allows for considerable levels of change in illumination and shape distortion.
5. **Keypoint matching:** These keypoints are matched between pairs of images of the chilli plant acquired from various angles and views. Bundle adjustment is used to form a sparse 3D point cloud of the plant and retrieve camera positions and intrinsic, extrinsic parameters simultaneously. The size, position, and orientation of the chilli plant are defined related to the coordinate frame of the reconstructed model.

Once the features are detected and matched from various views, a sparse 3D model is produced. The sparse model is filtered to remove outliers (erroneous peaks form due to keypoint mismatching) and unwanted reconstructed part is then removed manually. Thereafter, the calculated camera intrinsic and extrinsic parameters, positions, and orientations are used to generate a dense 3D point cloud of the plant. Cross-correlation matching method is used to match a pixel in one image with the corresponding pixel in an another image on the epipolar line for an overlapped image pair [23]. The process is repeated for each image pair.

The generated dense 3D point cloud then processed (cleaned, smoother, and managed) using remenshing and filtering tools in *Meshlab* [24] to create the final 3D model.

2.3 Investigation Parameters

This investigation considers two different parameters which have not been discussed yet in the literature.

1. **Change in number of input images for 3D modeling:** The 3D modeling method explained in Sect. 2.2 will be repeated on different subsets of randomly selected images for a particular set of input images. The size of the subset is varied from 25 images through to 78 images. For each subset size, the experiment is repeated five times, selecting a different random subset. The quality of the reconstructed 3D model will be determined by comparing features extracted from the model with ground truth data (manually measured values of the actual plant). Plant features, such as stem height and number of leaves will be extracted. By exploring the correlation of extracted features with ground truth values, the number of images required to give an accurate reconstruction of chilli plant will be determined.
2. **Change in image resolution:** Once the appropriate number of images for reconstruction is determined, the effects of image resolution on the 3D data will be analysed. This will provide information about the best image resolution for reconstruction of the chilli plant.

2.4 Measurement of Plant Features from the Reconstructed 3D Model

Two plant features: stem height and number of leaves are extracted manually from the 3D model. The stem height is calculated by calculating the distance between marked point on the stem tip and bottom in the reconstructed 3D model. These points are selected visually by the user. Figure 3 shows the marked points on the plant and the distance between the marked points. Similarly, the number of leaves can be estimated by zooming and rotating the reconstructed 3D model manually by the user.



Fig. 3 The approximate calculation of stem height from the reconstructed 3D model

3 Results and Discussion

3.1 3D Modeling

Detailed 3D modeling was performed as described in Sect. 2.2. The experiment on five different subsets of randomly selected images from the set of input images.

Using from 5 to 20 input images give poor 3D models. A small number of images do not cover enough plant views to provide sufficient images for an accurate reconstruction. Quality of the 3D model is directly proportional to the number of views and images [21]. Figure 4a shows a reconstructed 3D model using all 90 captured images. This model has replicated the actual plant very well when compared visually. Figure 4b shows the 3D model derived from 78 images. While there are some differences with Fig. 4a, these are minor. With fewer images, some of the details start to get lost, such as leaf stalks Fig. 4c and the stem Fig. 4d.

Consequently, features of the plant derived from the 3D model become less accurate. As leaves become disconnected from the stem, Fig. 4c, the number of leaves decreases, and as the stem becomes more broken (Fig. 4d–f), the measured stem height decreases. With fewer images, the number of points detected on the leaf surfaces decreases, distorting the shape and size of the leaves.

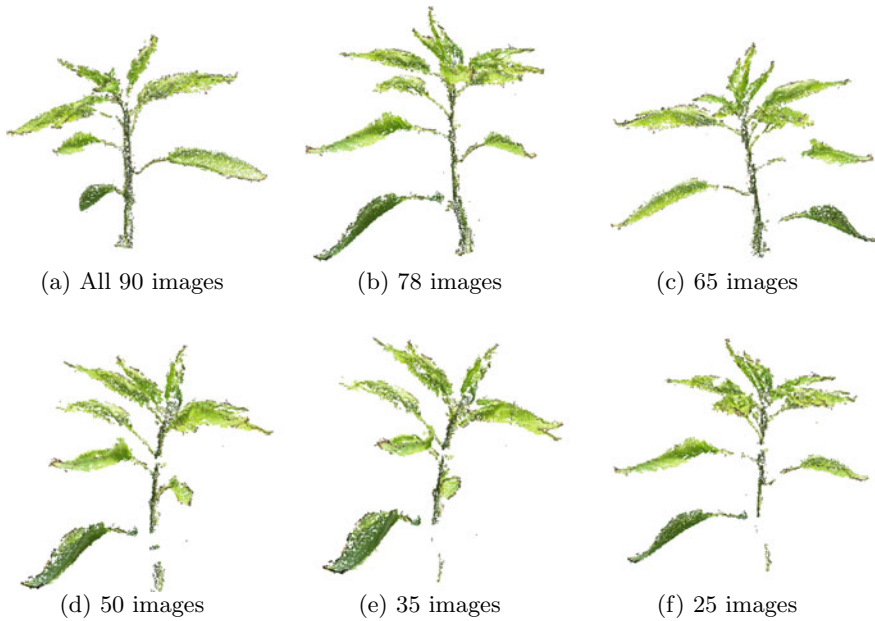


Fig. 4 Example 3D reconstructions of a chilli plant with different numbers of input images

3.2 Investigation Parameters

– Change in number of input images for 3D modeling:

For each subset size, the median, range, and interquartile range are represented by a box-whisker-plot. This clearly shows the variations from different random selections (see Fig. 5), and gives a clear idea of the accuracy of the extracted stem height for different numbers of input images. Similarly, Fig. 6 gives an idea about extracted number of leaves from 3D model with change in number of input images.

Using fewer images provide less information about plant features as they do not cover enough plant views. Due to insufficient plant views, the features from one image are not matched efficiently to another image. This causes feature matching error, resulting in unwanted outliers and loss of information. Because of the loss of information, the plant features, such as plant leaves are not reconstructed accurately and missing part of the leaf. Consequently, the leaf becomes disconnected from the stem so fewer leaves are counted on the reconstructed plant compared to the actual plant. Similarly, with fewer input images, the plant stem is not reconstructed accurately. The stem stem gives an inaccurate stem height from the reconstructed model.

Consequently, with more input images, the feature matching error and amount of information loss reduces as the input images cover sufficient plant views. It is important to use sufficient plant views to reconstruct the plant precisely. When input

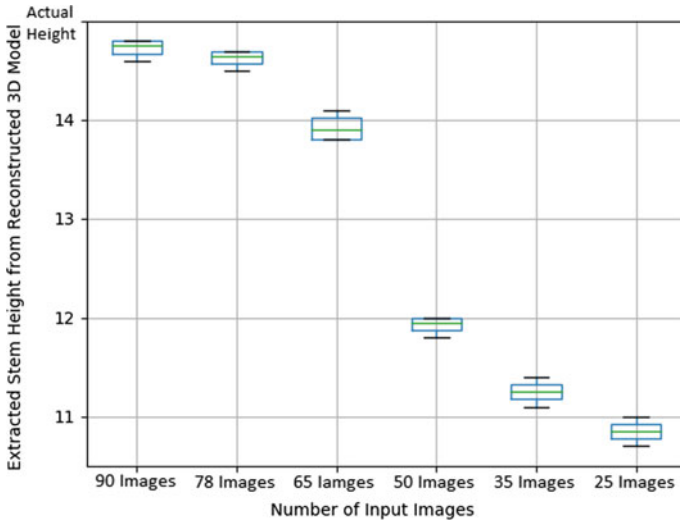


Fig. 5 Extracted stem height from reconstructed 3D model with change in number of input images

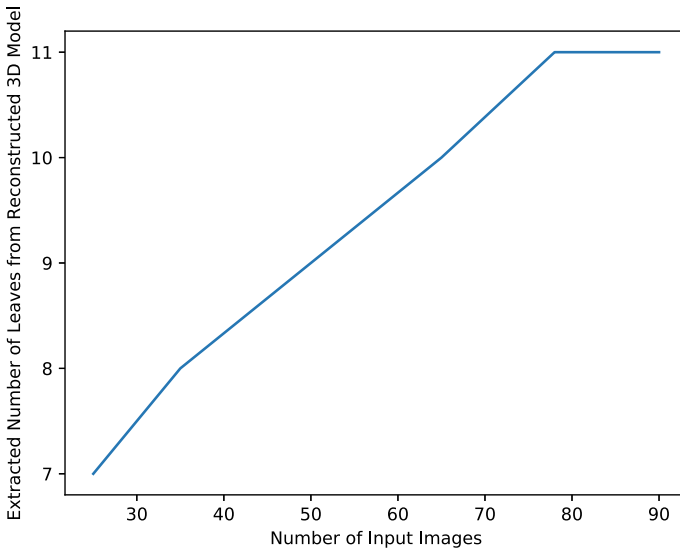
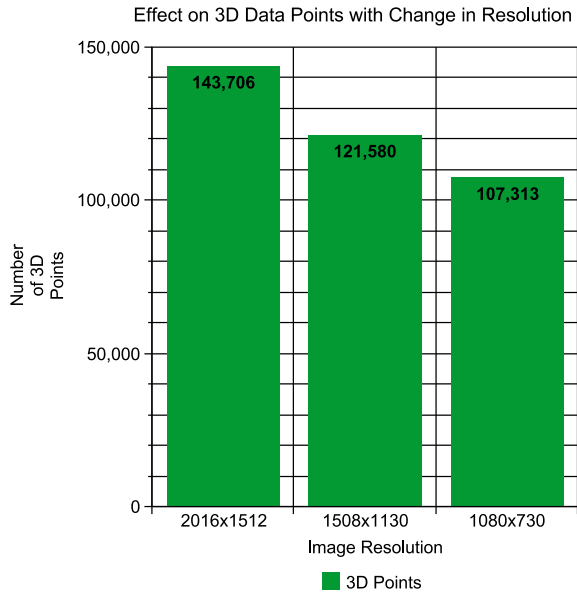


Fig. 6 Extracted number of leaves from reconstructed 3D model with change in number of input images

Fig. 7 Number of 3D data points detected for different image resolutions



images with sufficient plant views were considered, the error between extracted values from 3D model and ground truth values reduced. The difference between extracted values from 3D model and ground truth values varies inversely to number of input images. The 3D model with 78 input images provided good correlation between the extracted 3D values and ground truth values. From this experiment, it is clear that, instead of using all images, we can extract plant features precisely using fewer images which will inevitably save computation time and memory.

– Change in Image Resolution

The experiment in previous section has provided an approximate minimum number for accurate 3D reconstruction. In this section, the effect of changing image resolution will be explored. The original resolution of each image was 2016×1512 which gave 143,706 data points on average for the sample plant. These were sufficient to generate a precise 3D model. However, when the image size is reduced, there is a significant reduction in the number of 3D points detected, as shown in Fig. 7.

Although these data points are sufficient to reconstruct a 3D model, the quality of the model is poorer with fewer data points. When the resolution is further reduced, the number of data points also reduced to 107,313. The reduction in resolution tends result in inefficient feature detection and matching. Detecting fewer features leads to feature matching errors which are responsible to create unwanted outliers or distortions resulting in poor quality 3D model. Thus, the better the resolution of the images, the more 3D data points are detected.

4 Conclusion and Future Work

The results of experiments showed that 3D reconstruction of chilli plant and extracting plant features using fewer images is possible. However, the number of input images required for precise 3D reconstruction is totally depends on plant architecture. If the plant architecture is complex then more number of input images will be required for 3D reconstruction. Similarly, if plant architecture is simple then fewer images are sufficient for precise 3D reconstruction.

Our contribution to the knowledge is: (1) Investigation of the effect on 3D model when there is a change in the number of input images. (2) Guidelines for the selection of an appropriate number of images for accurate 3D resolution. (3) Successful extraction of plant features non-destructively from a 3D model non-destructively. (4) Investigation of the effect on 3D data when there is change in image resolution. The 3D model reconstructed from fewer images can provide more information about plant like leaf area, leaf length, leaf width, stem diameter and so on. The future work can be extended to implement this method on more plants and analyse the required input images and image resolutions. In addition, it is still needed to analyse, if we carefully select fewer images manually, is it possible to reduce the number of input images and still generate a precise 3D model?

References

1. Paturkar A, Gupta GS, Bailey D (2019) 3D reconstruction of plants under outdoor conditions using image-based computer vision. In: Santosh KC, Hegadi RS (eds) Recent trends in image processing and pattern recognition. Springer, Singapore, pp 284–297
2. Mishra KB, Mishra A, Klem K, Govindjee (2016) Plant phenotyping: a perspective. *Ind J Plant Physiol* 21(4):514–527 (2016)
3. Li L, Zhang Q, Huang D (2014) A review of imaging techniques for plant phenotyping. *Sensors* 14(11):20 078–20 111
4. Paturkar A, Gupta GS, Bailey D (2017) Overview of image-based 3D vision systems for agricultural applications. In: International conference on image and vision computing New Zealand (IVCNZ), pp 1–6
5. Guo Q, Wu F, Pang S, Zhao X, Chen L, Liu J, Xue B, Xu G, Li L, Jing H, Chu C (2018) Crop 3D—A LiDAR based platform for 3D high-throughput crop phenotyping. *Sci China Life Sci* 61(3):328–339
6. Zhang Z (2012) Microsoft kinect sensor and its effect. *IEEE MultiMedia* 19(2):4–10
7. Scharstein D, Szeliski R (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int J Comput Vis* 47(1–3):7–42
8. Jebara T, Azarbajejani A, Pentland A (1999) 3D structure from 2D motion. *IEEE Signal Process Mag* 16(3):66–84
9. Paulus S, Dupuis J, Riedel S, Kuhlmann H (2014) Automated analysis of barley organs using 3D laser scanning: an approach for high throughput phenotyping. *Sensors* 14(7):12 670–12 686
10. Kaminuma E, Heida N, Tsumoto Y, Yamamoto N, Goto N, Okamoto N, Konagaya A, Matsui M, Toyoda T (2004) Automatic quantification of morphological traits via three-dimensional measurement of Arabidopsis. *Plant J* 38(2):358–365

11. Newcombe RA, Izadi S, Hilliges O, Molyneaux D, Kim D, Davison AJ, Kohi P, Shotton J, Hodges S, Fitzgibbon A (2011) Kinectfusion: real-time dense surface mapping and tracking. In: 10th IEEE international symposium on mixed and augmented reality (ISMAR), pp 127–136
12. Chéné Y, Rousseau D, Lucidarme P, Bertheloot J, Caffier V, Morel P, Belin T, Chapeau-Blondeau E (2012) On the use of depth camera for 3D phenotyping of entire plants. *Comput Electron Agric* 82:122 – 127
13. Baumberg A, Lyons A, Taylor R (2005) 3D S.O.M.A commercial software solution to 3D scanning. *Graph Models* 67(6):476–495
14. Bergerman M, Van Henten E, Billingsley J, Reid J, Mingcong D (2013) IEEE robotics and automation society technical committee on agricultural robotics and automation. *IEEE Robot Autom Mag* 20(2):20–23
15. Takizawa H, Yamamoto S, Ezaki N, Mizuno S (2005) Plant recognition by integrating color and range data obtained through stereo vision. *J Adv Comput Intell Intell Inform*
16. Ivanov N, Boissard P, Chapron M, Andrieu B (1995) Computer stereo plotting for 3-D reconstruction of a maize canopy. *Agric Forest Meteorol* 75(1):85–102
17. Jay S, Rabatel G, Hadoux X, Moura D, Gorretta N (2015) In-field crop row phenotyping from 3D modeling performed using structure from motion. *Comput Electron Agric* 110:70–77
18. Quan L, Tan P, Zeng G, Yuan L, Wang J, Kang SB (2006) Image-based plant modeling. *ACM Trans Graph* 25(3):599–604
19. Tan P, Zeng G, Wang J, Kang SB, Quan L (2007) Image-based tree modeling. *ACM Trans Graph* 26(3):87
20. Hu Y, Wang L, Xiang L, Wu Q, Jiang H (2018) Automatic non-destructive growth measurement of leafy vegetables based on kinect. *Sensors* 18:806
21. Liu S-X, An P, Zhang Z-Y, Zhang Q, Shen L-Q, Jiang G-Y (2009) On the relationship between multi-view data capturing and quality of rendered virtual view. *Imaging Sci J* 57(5):250–259
22. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110
23. Remondino F, El-Hakim S (2005) Critical overview of image-based 3D modeling. In: International workshop on recording, modeling and visualization of cultural heritage. CRC Press, p 299
24. Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli E, Ranzuglia G (2008) Meshlab: an open-source mesh processing tool. In: Scarano V, Chiara RD, Erra U (eds) Eurographics Italian chapter conference. The Eurographics Association