3D Reconstruction of Plants Under Outdoor Conditions Using Image-Based Computer Vision

Abhipray Paturkar$^{(\boxtimes)}$, Gaurab Sen Gupta, and Donald Bailey

School of Engineering and Advanced Technology, Massey University,
Palmerston North, New Zealand
A.Paturkar@massey.ac.nz

Abstract. 3D reconstruction of plants under outdoor conditions is a challenging task, for applications such as plant phenotyping which needs non-invasive methods. With the availability of new sensors and reconstructions techniques, 3D reconstruction is improving rapidly. However, sensors are still expensive for researchers. In this paper, we propose a cost-effective image-based 3D reconstruction approach which can be achieved by off-the-shelf cameras. This approach is based on the structure-from-motion method. We implemented this approach in MATLAB and Meshlab is used for further processing to achieve an exact 3D model. We also investigated the effect of different adverse outdoor scenarios which affect quality of 3D model such as movement of plants because of strong wind, drastic change in light condition while capturing the images. We have decreased the appropriate number of images needed to get precise 3D model. This method gives accurate results and it is a fast platform for non-invasive plant phenotyping.

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

1 Introduction

In precision agriculture, plant phenotyping is an important aspect. It helps scientists and researchers to collect valuable information regarding plant structure, which is inevitably a basic requirement to enhance plant discrimination and plant selection [1]. 3D reconstruction models of the plant through phenotyping operations are helpful for evaluating plant growth and yield over time. This permits management of the plant to be more extensive [2]. This 3D reconstructed plant models could be used to describe leaf features, discriminate between weed and crop, estimate the biomass of the plant, and classify fruits. Conventionally all these elements have been evaluated by experts in this field depending on the visual score, which was responsible for creating dissimilarity between expert judgements. In addition, this process is tedious.

Primarily, the aim of plant phenotyping is to calculate plant features precisely without subjective biases. Nonetheless, developing expertise and knowledge still lack technical advancements in processing and sensing technologies. Most of the modern sensing technologies are primarily only two dimensional, e.g. thermal or hyperspectral imaging. Inferring 3D information from such sensors is greatly reliant on distance and angle to the plants. In contrast, 3D reconstruction is being suggested for morphological classification of plants. 3D reconstruction is developing rapidly and getting tremendous attention. Structured light (Kinect sensor) [3], ToF cameras [4] and LiDAR [5] are active sensing techniques used for 3D reconstruction which basically use their own source of illumination. However, these state-of-the-art systems are costly. On the other hand, image-based passive 3D reconstruction techniques which use radiation present in the scene, which includes, structure-from-motion [6], stereo vision [7] and space carving system [8], only need one or two cameras which results in a very cost-effective system.

ToF cameras have excellent performance and appeared to be a suitable sensor for evaluating plants. ToF cameras are commonly combined with an RGB camera. Kazmi et al. [4] analysed the performance of ToF cameras for close range imaging in different illumination conditions. They found that ToF cameras deliver high frame rates as well as accurate depth data under suitable conditions. However resolutions of depth images are often low; the sensors are sensitive to ambient sunlight, which usually leads to poor performance while working outdoors; the quality of depth values depends on the color of objects, and some sensors have blurring problems while sensing moving objects. Because of these limitations, it is difficult to use TOF cameras for 3D reconstruction under outdoor conditions.

LiDAR is an expansion of the principles applied in radar technology. It estimates the distance between the target and the scanner by illuminating the target using a laser and calculating the time taken for the reflected light to come back [9]. Kaminuma et al. [10] presented an application of a laser range finder for 3D reconstruction which represents the leaves and as a polygonal meshes and then measured the morphological features from those models. Paulus et al. [11] determined that LiDAR is an appropriate sensor for obtaining precise 3D point clouds of plants but it does not give any information on the surface area. In addition, it had poor resolution and a long warm-up time. In contrast, LiDAR has given excellent results under outdoor conditions having a drawback of being very costly. Other disadvantages of the LiDAR sensor are that it needs calibration and multiple captures are required to overcome issues with occlusion. The data from the sensors cannot detect leaves overlapping efficiently and depth and images are not of high quality.

An alternative approach for depth estimation is the use of structured light. In this approach, the light source (either near-infrared or visible) is offset a familiar distance from an imaging device. The luminous from the emitter is reflected into the camera by the target object. Information about the light pattern allows the depth to be derived through triangulation [9]. Baumberg et al. [12] presented a 3D plant analysis based on the technique they called mesh processing. In this work,

the authors made a 3D reconstructed model of a cotton using Kinect sensor which performed well under indoor conditions yet struggled under outdoor conditions. Chéné et al. [13] used a depth camera to segment plant leaves and reconstructed the plant in 3D.

As mentioned above, stereo vision and structure-from-motion use passive illumination, which allows these techniques to work efficiently under outdoor conditions. A off-the-shelf digital camera could be used for capturing overlapped images which are processed by a computer to estimate the depth or 3D reconstructed model. Stereo vision has comparatively lower cost than active sensing techniques and has provided excellent 3D reconstructed models. Nevertheless, the camera alignment and spacing between the cameras should be precise. As an illustration, the distance between the plant and camera, which is calculated with the help of the focal length of the camera, there should be an overlapping between the images and the rotation of the plant in different images. Ivanov et al. [14] described maize plants under outdoor conditions by using images captured from various angles to characterize plant structure. Takizawa et al. [15] reconstructed a 3D model of a plant and derived plant height and shape information.

The combination of images and cameras in structure-from-motion generally create a sparse 3D point cloud. Structure-from-motion consists of calculating a set of points from position of cameras, from these set of points, a dense point cloud is created. Jay et al. [16] proposed a method which builds a 3D reconstructed model of a crop row to get the plant structural parameters. This 3D model is acquired using structure-from-motion with the help of colour images captured by translating a single camera along the row. Quan et al. [17,18] proposed a semi-automatic method for modelling plants for application like plant phenotyping, yield estimation based on structure-from-motion which performed well under outdoor conditions but it is computationally expensive.

In summary, each sensing technique has some merits and demerits [9]. The need of current sensors and systems is to reduce the need for manual extraction of phenotypic data. Their performance stays, to a lesser or greater extent, restricted by the dynamic morphological complications of plants [19]. Currently, there is not a 3D system and method which solves all necessities, but one should select depending on the budget and requirements. Moreover, plant structure is generally complex which includes a large amount of self-occlusion (leaves blocking one another). Hence, reconstructing plants in 3D in non-invasive manner stays a serious challenge in phenotyping.

Focusing at contributing a cost-effective solution to above challenge, we present an image-based 3D reconstruction system under outdoor conditions. Our contributions include:

1. An easy and cost-effective system (using just a mobile phone camera)
2. Investigation of effects of adverse outdoor scenarios and possible solutions (movement of plants because of wind and change in light condition because of the movement of clouds while capturing the images)
3. A precise 3D model obtained from a limited number of images

The rest of the paper is organized as follows; Sect. 2 discusses the step by step results of the method we used in this paper. The effect of adverse outdoor scenarios on 3D model along with the possible solution discussed in Sect. 3.

2   Materials and Methods

We selected a chilli plant *Capsicum annum L.* on a commercial field (Palmerston North, New Zealand) for testing our image processing. This chilli plant is selected for its demand over the year and its high value. Images were acquired during December 2017 when plant height was between 15 cm to 20 cm. A crop
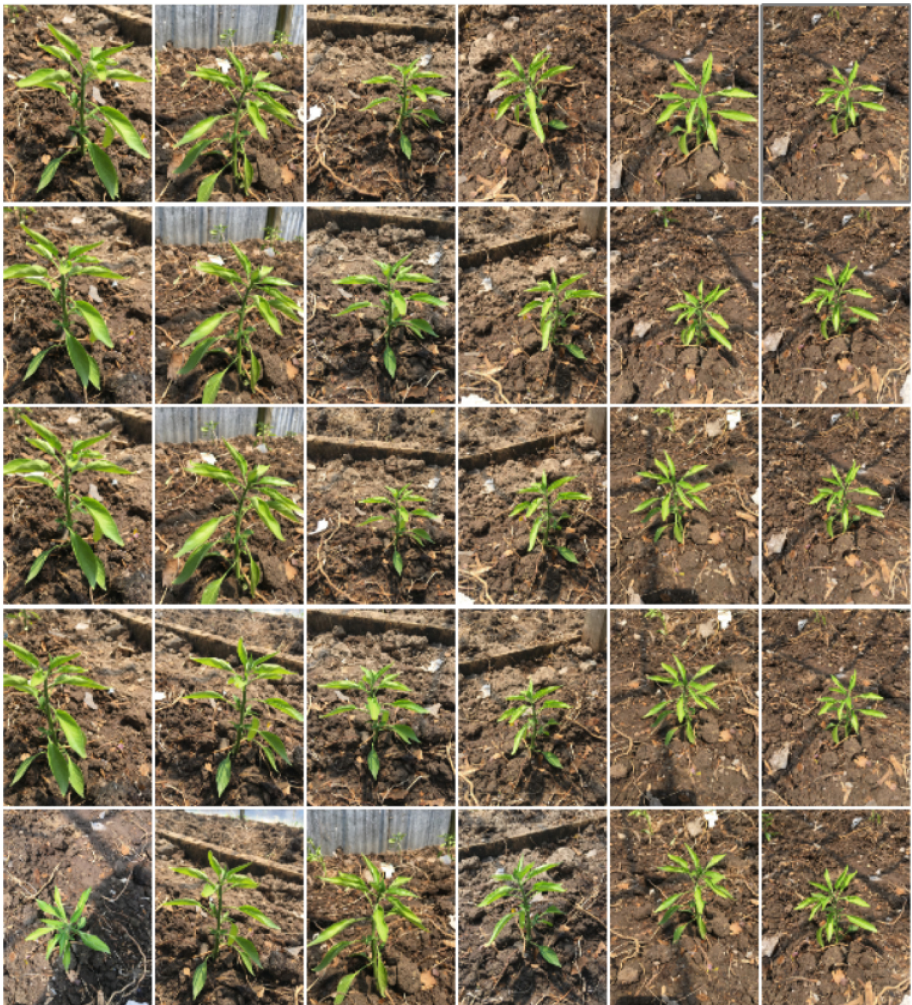

Fig. 1. Sample of the captured images with different view angles of a chilli plant

was planted in lines 90 cm apart, our experiment aimed at modelling individual plants. As a result, other plants did not hinder in the model and only one plant was monitored at a time.

2.1   Image Acquisition

The images were captured sequentially following a circular path with respect to the plant axis. Seven different rounds were taken at various angles, heights and distance. At least 15 images were captured at each path by revolving around the plant with a mobile phone's rear camera (Apple iPhone 6s+ with 12MP rear camera, f/2.2), capturing at every 10° to 15° of the perimeter. The distance between plant and camera was not kept constant. These seven rounds made during the image acquisition process produced 105 images with 95% overlap between successive images. Images were taken under outdoor conditions. This gives us variety of images to work with. The camera positions were chosen to ensure that the plant was entirely in the field of view, and the images were of a good quality (not blurred etc.). Structure-from-motion calculates the intrinsic camera parameters by itself, so camera positions do not have to be calibrated during the image acquisition process. Samples of the captured images with different view angles of chilli plant and image acquisition scheme is shown in Figs 1 and 2 respectively.

2.2   Plant-Soil Segmentation

As we are conducting our experiment under outdoor conditions, plant-soil segmentation has to be robust. This step is to distinguish plant-pixels from soil-
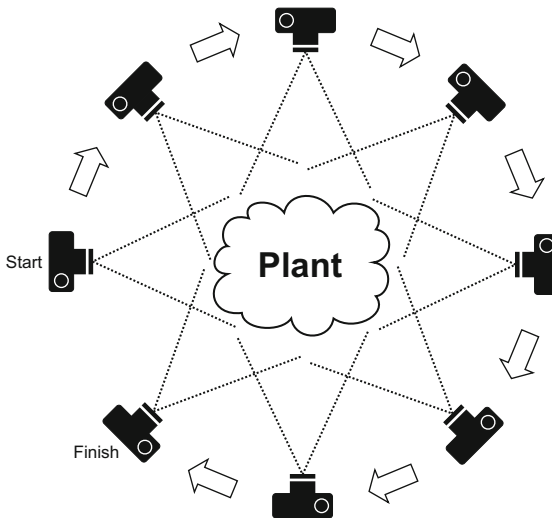

Fig. 2. Image acquisition scheme

pixels. As this process is applied to every image, this segmentation has to be autonomous. The improved vegetation index, excess green (ExG) [20] has been used, which is defined as:

$$ExG = \frac{2G - R - B}{R + G + B} \tag{1}$$

where, R, G, and B are the red, green, and blue pixel components respectively. With ExG, pixels associated with the plant class generally have high ExG values. This makes the discrimination between plant and soil easier. Figure 3 depicts plant-soil segmentation of one of the views.

Fig. 3. Plant-soil segmentation based on improved vegetation index

2.3   Keypoint Detection and Matching

After segmentation between plant and soil, the next task is to find the common keypoints (features) between a pair of images. For this process, we implemented the scale-invariant feature transform (SIFT) [21]. We converted an image into a huge set of keypoint vector, all of them is invariant to image scaling, rotation and translation. The standard steps in SIFT are.

1. **Formation of a scale space:** A basic step of calculation explores over each and every scales and image locations. It is achieved decisively using difference-of-Gaussian (DoG) function to determine potential keypoints that are scale and orientation invariant.

2. **Locating keypoints:** As we located the possible keypoints, a structured model is fit to identify scale and location. These keypoints are chosen hinged on their stability.
3. **Assignment of orientation:** Depend on local image gradient directions, one or additional orientations are elected to the location of every keypoint. Each and every operations are executed on image data which has been transformed corresponding to the elected scale, location, and orientation for each keypoint, by that giving invariance to all these transformations.
4. **Keypoint descriptor:** In the preferred scale in the region near to every keypoint, the local image gradients are calculated. There gradients are transformed into a delineation that permits for considerable levels of change in illumination and local shape distortion. Figure 4 illustrate the keypoints detected in two images.
5. **Matching of keypoints:** Keypoints are matched between pair of images of an object or a scene captured from different view points and angles. Matching is based on finding similar keypoint feature vectors between the two images. Figure 5 shows matching keypoints between two images. The matches are then filtered to remove outliers, and bundle adjustment is used to create a sparse 3D point cloud of matching object or scene and to retrieve camera calibration intrinsic, extrinsic parameters and positions at the same time. Pyramid-like symbol in Fig. 6 represents the positions and angles of the camera and green dots represent the plant structure.
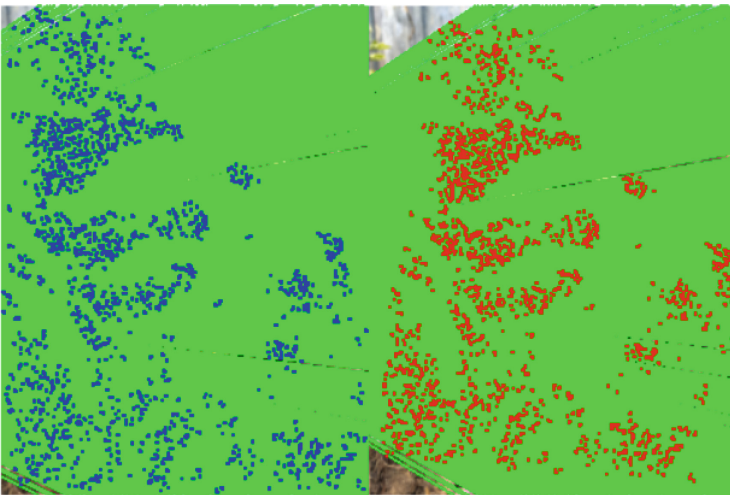

Fig. 4. Keypoint detection in an image pair

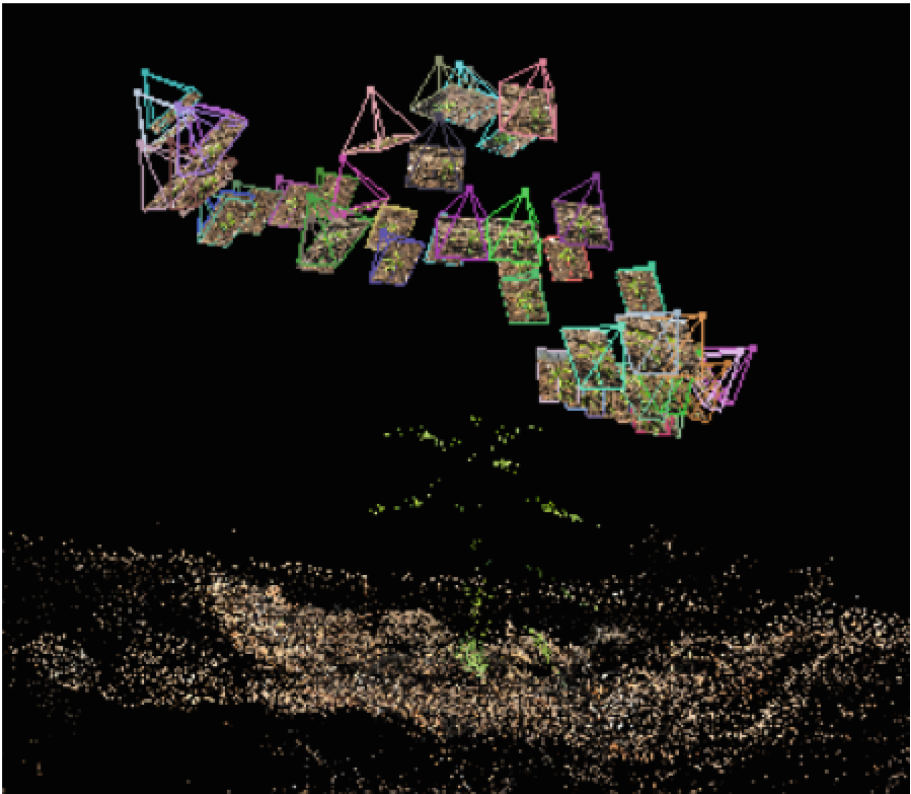**Fig. 5.** Keypoint matching in an image pair

Fig. 6. Sparse 3D point cloud of the scene

2.4   3D Reconstruction

Finally, the calculated camera positions, parameters, and orientations are used to create a dense 3D point cloud. We implemented a cross-correlation matching method. For a pair of overlapped images, a pixel in the first image is corresponded with the pixel corresponding in the second image on the epipolar line [7]. This process is iterated for each pair of images keeping in mind that the calculated position of a given keypoint to be less noisy. The derived dense 3D point cloud is shown in Fig. 7, because of the page limitations we have added just two views as a resultant 3D model.

Fig. 7. 3D reconstructed model of chilli plant

2.5   Post Processing

The dense 3D point cloud is post processed off-line in an open source software named Meshlab [22]. This software is used to process unstructured dense 3D models using filters and remeshing tools, which helps to clean, smooth and manage our dense 3D model, which helps us to solve the quantization issue. Figure 8 shows the cleaned entire 3D model.

2.6   Selection of Appropriate Number of Images

It is very tricky to decide the number of images needed for plant 3D reconstruction, and hence it is an important factor. In general, a larger number of images will give additional information about the plant. At the same time, it will hold redundant data because of the overlapping regions of same scene, and it will take extra computation time to process more images. Moreover, it was noticed during our experiment that a large number of images caused feature matching error
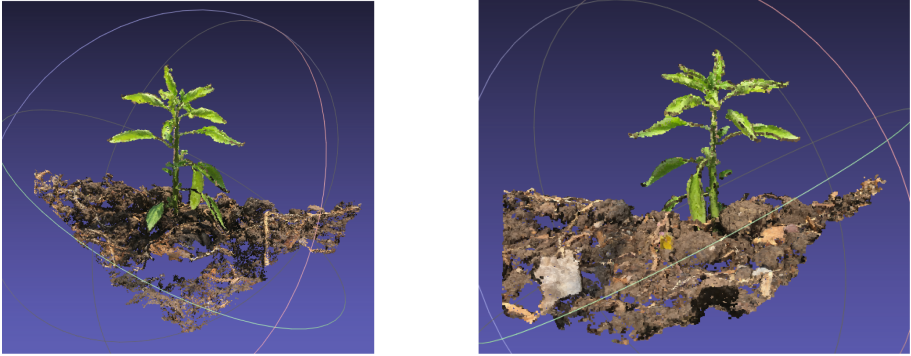
Fig. 8. Clean 3D reconstructed model with overlapping leaves

which inevitably affects the accuracy. In contrast, with few images, the output 3D model will lack necessary data about the plant. We determined during our experimentation that it is quite difficult to reconstruct the plant in 3D using just 3–4 images, which cover only a limited range of viewpoints.

So based on our above investigation, hypothesis around the connection between multi-view information capturing and the trait of interpreted virtual view were tested to find an appropriate balance between multi-view information capturing and the quality of the 3D reconstructed model [23].
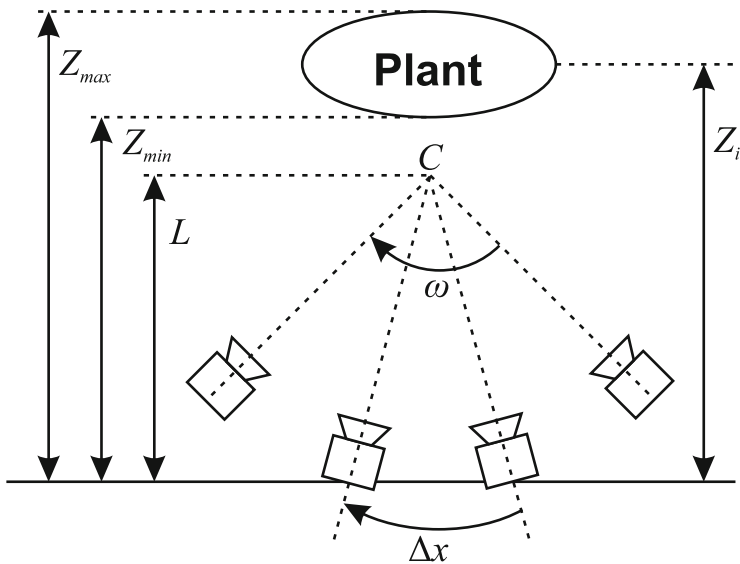

Fig. 9. Camera model

Figure 9 illustrates the camera model used in this experiment. Zi is the distance between plant and camera. ω is an arc which shows the space between each view from the camera with radius L in same pitch which is $= \Delta x$. f_l is a focal length of the camera. Z_{max} is the maximum depth of the plant and Z_{min} is the minimum depth of the plant.

Based on these assumptions and model, an appropriate number of images for 3D reconstruction can be calculated based on the below formulas:

$$\Delta x_{max} = \frac{1}{f_l} \frac{Z_{max}}{Z_{min}} \tag{2}$$

$$f_{nyq} = \frac{L}{2\Delta x_{max}} \tag{3}$$

where

$$L = \frac{2 \times Z_{max} Z_{min}}{Z_{max} + Z_{min}} \tag{4}$$

$$N_A = \frac{\omega L}{\Delta x_{max}} \tag{5}$$

We selected 30 as an appropriate image number for 3D reconstruction based on the above theory and formulas. Due to the page limitations we are not presenting the step by step calculation of the aforementioned formulas but as it is straight forward theory, it is easy to estimate the appropriate number of images.

3   Discussion

The step by step results of the experimentation have been shown throughout the paper. It is difficult to quantify the quality of the 3D model but according to the rule of thumb, quality of the 3D model is a function of its input size to realism it produces. So to validate our result, visual analysis of the 3D models we achieved (Fig. 8) are compared with the result presented in [24], which illustrate that our 3D models are having better quality as our models are not missing any details like petioles, surface of the leaves and flower buds. There are different validation approaches given in literature. Several studies involved the extraction of 2D visual records and compared to measurements achieved by manual phenotyping. Another approach is to use the different databases that have allowed researchers to assess the accuracy of their 3D model [8,25].

In this experimentation, we captured numerous images of the chilli plant and the number is ranged from 5 to 100 images. We selected 30 as an appropriate image number according to the theory presented and the quality of 3D model.

3.1   Effect of Adverse Outdoor Scenarios:

Based on our literature survey and our outdoor experimentations, we analysed that there are still some scenarios which cause problems and need more attention. In general, we know that sometimes under outdoor conditions it can be

windy. We acquired another set of images of plants in windy condition, where
plants were moving. In another scenario we captured the images when there was
change in light conditions because of movement of the clouds. Here, we tried to
investigate the effect of these scenarios under outdoor conditions and the effects
of these on the resulting 3D models.

Fig. 10. Poor feature matching and 3D reconstruction due to heavy wind

Movement of Plant: In this scenario, we noticed that, because of the displace-
ment of the plant due to wind, there were many feature matching errors resulting
in poor 3D model. The resulting 3D model was missing important details in the
stem area of the plant with some half reconstructed leaves (see Fig. 10). One
possible solution for this scenario is to detect the inconsistent matches between
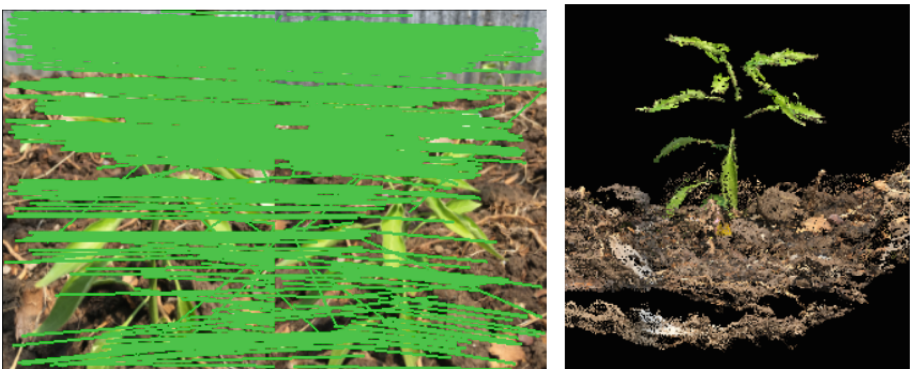the images because of the wind and filter out those images from the database.

Fig. 11. Poor feature matching and 3D reconstruction due to drastic change in illumi-
nation

Change in Illumination: In our study, we try to reduce the error caused by change in illumination, with good results. However, in certain scenarios, there could be a drastic change in the illumination while capturing the plant images. We studied that, in this scenario, the resulting 3D model was missing necessary information about the plant such as plant surface and leaves resulting in blank patches in the 3D model, shown in Fig. 11. One possible solution for this scenario is to pre-process and normalise the acquired images first to reduce the effect of change in illumination in the database.

4   Conclusion

Plant phenotyping is achievable using our approach. Arguably, results of our experiments demonstrated that chilli plant 3D reconstruction is feasible with a low budget and could be used in different scenarios, even under outdoor conditions. Our contribution contains: (1) Easy and cost-effective system operated under outdoor conditions and achieved good results. (2) Investigation of adverse outdoor scenarios and effect on 3D model (3) The appropriate number of images were selected and used for reconstruction. (3) An entire 3D model with limited images. (4) Automatic plant-soil segmentation is implemented. This 3D reconstruction system is gives a cost-effective and efficient platform for non-invasive plant phenotyping, containing informations such as, fruit volume, leaf angles, leaf area index, which are important for assessing the stress and growth on plant features.

References

1. Mishra, K.B., Mishra, A., Klem, K., Govindjee: Plant phenotyping: a perspective. Indian J. Plant Physiol. **21**(4), 514–527 (2016)
2. Li, L., Zhang, Q., Huang, D.: A review of imaging techniques for plant phenotyping. Sensors **14**(11), 20078–20111 (2014)
3. Zhang, Z.: Microsoft Kinect sensor and its effect. IEEE Multimedia **19**(2), 4–10 (2012)
4. Kazmi, W., Foix, S., Alenyà, G., Andersen, H.J.: Indoor and outdoor depth imaging of leaves with time-of-flight and stereo vision sensors: analysis and comparison. ISPRS J. Photogramm. Remote Sens. **88**, 128–146 (2014)
5. Guo, Q., et al.: Crop 3D—a LiDAR based platform for 3D high-throughput crop phenotyping. Sci. China Life Sci. **61**(3), 328–339 (2018)
6. Jebara, T., Azarbayejani, A., Pentland, A.: 3D structure from 2D motion. IEEE Signal Process. Mag. **16**(3), 66–84 (1999)
7. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vision **47**(1–3), 7–42 (2002)
8. Cremers, D., Kolev, K.: Multiview stereo and silhouette consistency via convex functionals over convex domains. IEEE Trans. Pattern Anal. Mach. Intell. **33**(6), 1161–1174 (2011)
9. Paturkar, A., Gupta, G.S., Bailey, D.: Overview of image-based 3D vision systems for agricultural applications. In: 2017 International Conference on Image and Vision Computing New Zealand (IVCNZ), pp. 1–6, December 2017

10. Kaminuma, E., et al.: Automatic quantification of morphological traits via three-dimensional measurement of Arabidopsis. Plant J. **38**(2), 358–365 (2004)
11. Paulus, S., Dupuis, J., Riedel, S., Kuhlmann, H.: Automated analysis of barley organs using 3D laser scanning: an approach for high throughput phenotyping. Sensors **14**(7), 12670–12686 (2014)
12. Baumberg, A., Lyons, A., Taylor, R.: 3D S.O.M.—a commercial software solution to 3D scanning. Graph. Models **67**(6), 476–495 (2005)
13. Chéné, Y., et al.: On the use of depth camera for 3D phenotyping of entire plants. Comput. Electron. Agric. **82**, 122–127 (2012)
14. Ivanov, N., Boissard, P., Chapron, M., Andrieu, B.: Computer stereo plotting for 3-D reconstruction of a maize canopy. Agric. For. Meteorol. **75**(1), 85–102 (1995)
15. Takizawa, H., Yamamoto, S., Ezaki, N., Mizuno, S.: Plant recognition by integrating color and range data obtained through stereo vision. J. Adv. Comput. Intell. Intell. Inform. **9**(6), 630–636 (2005)
16. Jay, S., Rabatel, G., Hadoux, X., Moura, D., Gorretta, N.: In-field crop row phenotyping from 3D modeling performed using structure from motion. Comput. Electron. Agric. **110**, 70–77 (2015)
17. Quan, L., Tan, P., Zeng, G., Yuan, L., Wang, J., Kang, S.B.: Image-based plant modeling. ACM Trans. Graph. **25**(3), 599–604 (2006)
18. Tan, P., Zeng, G., Wang, J., Kang, S.B., Quan, L.: Image-based tree modeling. ACM Trans. Graph. **26**(3), 87 (2007)
19. Paproki, A., Sirault, X., Berry, S., Furbank, R., Fripp, J.: A novel mesh processing based technique for 3D plant analysis. BMC Plant Biol. **12**(1), 63 (2012)
20. Meyer, G., Camargo Neto, J.: Verification of color vegetation indices for automated crop imaging applications. Comput. Electron. Agric. **63**, 282–293 (2008)
21. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004)
22. Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G.: Meshlab: an open-source mesh processing tool. In: Scarano, V., Chiara, R.D., Erra, U. (eds.) Eurographics Italian Chapter Conference. The Eurographics Association (2008)
23. Liu, S.-X., An, P., Zhang, Z.-Y., Zhang, Q., Shen, L.-Q., Jiang, G.-Y.: On the relationship between multi-view data capturing and quality of rendered virtual view. Imaging Sci. J. **57**(5), 250–259 (2009)
24. Ni, Z., Burks, T., Lee, W.: 3D reconstruction of plant/tree canopy using monocular and binocular vision. J. Imaging **2**(4), 28 (2016)
25. Pound, M.P., French, A.P., Murchie, E.H., Pridmore, T.P.: Automated recovery of three-dimensional models of plant shoots from multiple color images. Plant Physiol. **166**(4), 1688–1698 (2014)