Accurate Multi-View Stereo 3D Reconstruction for Cost-Effective Plant Phenotyping

Lu Lou[1](✉), Yonghuai Liu[1], Jiwan Han[2], and John H. Doonan[2]

[1] Department of Computer Science, Aberystwyth University, Aberystwyth, UK
[2] NPPC, IBERS, Aberystwyth University, Aberystwyth, UK
lul1@aber.ac.uk

Abstract. Phenotyping, which underpins much of plant biology and breeding, involves the measurement of characteristics or traits. Traditionally, this has been often destructive and/or subjective but the dynamic objective measurement of traits as they change in response to genetic mutation or environmental influences is an important goal. 3-D imaging technologies are increasingly incorporated into mass produced consumer goods (3D laser scanning, structured light and digital photography) and may represent a cost-effective alternative to current commercial phenotyping platforms. We evaluate their performance, cost and practicability for plant phenotyping and present a 3D reconstruction method for plants from multi-view images acquired with domestic quality cameras. We exploit an efficient Structure-From-Motion followed by stereo matching and depth-map merging processes. Experimental results show that the proposed method is flexible, adaptable and inexpensive, and promising as an generalized groundwork for phenotyping various plant species.

Keywords: Plant phenotyping · Multi-view images · Structure from motion · Stereovision · 3D reconstruction

1 Introduction

The phenotype of an organism emerges from the interaction of its genotype with its developmental history and its environment. The range of phenotypes, therefore can be large, even from a single genotype. The measurement of phenotypes, as they change in response to genetic mutation and environmental influences, is a laborious and expensive process. Phenotyping is a major bottleneck that limits the exploitation of genomics in plant science, as well as animal biology and medicine. Important requirements include improving the accuracy (how close the measurement is to the absolute ground truth), precision (the repeatability or variance of the measurement process) and throughput of phenotyping at all levels of biological organization while reducing costs and minimizing human labor by means of automation, integrated techniques and experimental design [2].

A. Campilho and M. Kamel (Eds.): ICIAR 2014, Part II, LNCS 8815, pp. 349–356, 2014.
DOI: 10.1007/978-3-319-11755-3_39

In recent years, various automatic high-throughput plant growth and phenotyping platforms have been developed, PHENOPSIS [4] is used by French National Institute for Agricultural Research (INRA) for Arabidopsis. TraitMillTM [15], developed by the company CropDesign, was used to evaluate transgenic rice (Oryza sativa). Commercial high-throughput phenotyping platforms, with automated plant handling and imaging systems that capture morphological and physiological data using a variety of sensors, are aimed at reducing the need for manual acquisition of phenotypic data but do so at considerable cost. Their functionality remains, to a greater or lesser extent, limited by the dynamic morphological complexity of plants.

Although commencing as a relatively simple structure, the plant body plan rapidly becomes more complex due to a variety of processes – re-iterative organ formation, changes in organ spacing and identity, branching – that lead to overlapping and variable 3-D organization. Recording and measuring this complexity in a dynamic (non destructive) manner remains a serious challenge in biology.

The extent of this challenge is illustrated by the diversity of proposed methods that include laser scanning, digital camera photography, and structured light ranging. Stereo vision was used to reconstruct the 3D surface of maize for measurement and analysis [7]. Kaminumaet et al., [9] applied a laser range finder to reconstruct 3D models that represented the leaves and petioles as polygonal meshes and then quantified morphological traits from these models. Biskup designed a stereo vision system with two cameras to build 3D models of soybean plants foliage and analyzed the angle of inclination of the leaves and its movement over time [1]. Quan proposed a method to interactively create a 3D model of the foliage by combining clustering, image segmentation and polygonal models [14]. 3D plant analysis based on mesh processing technique was presented in [13], where the authors created a 3D model of cotton from high-resolution images using a commercial 3D digitization product named 3DSOM. Thiago developed an image-based 3D digitizing method for plant architecture analysis and phenotyping [17], and showed that the state-of-the-art SFM and multi-view stereovision are able to produce accurate 3D models with a few limitations. Li et al. [10] presented a framework to track and detect plant growth by a forward-backward 3D point cloud analysis, where the 3D point cloud was produced based on a structured light scanner over time. Yamazaki et al. presented a practical shape-from-silhouettes approach to acquire 3D models of intricate objects with severe self-occlusions, repeated thin structures, and surface discontinuities, including tree branches, bicycles and insects [18].

In this paper, we describe an accurate multi-view stereo (MVS) 3D reconstruction method of plants using multi-view images, which takes both accuracy and efficiency into account. The key of our method is efficient Structure-From-Motion followed by a stereo matching and depth-map merging process. Compared to the state-of-art MVS methods, the proposed method has three main advantages: i) It uses structure-from-motion to estimate the camera's parameters from an sequence of images; ii) It can reconstruct accurate and dense point clouds; iii) It is computationally efficient and therefore much faster than other state-of-the-art methods.

2 The Proposed Method

3D sensors being developed for non-destructive plant phenotyping are based, variously, on laser scanning , structured light, multi-view stereovision, visual hull, etc., and each has its own merits and limitations. Unfortunately, there does not exist a widely accepted benchmark that can be used to compare those different methods. Therefore, we have investigated the existing methods of 3D reconstruction and conclude that the 3D laser/lidar scanner (we did not consider the hand-held laser scanner due to its low efficiency) or the structured-light scanner (including Kinect sensor) do not work well on plants, especially on complex or even marginally occluded specimens or tiny plants. Consequently, we sought to build an efficient and accurate image-based 3D reconstruction system that could cope with a diversity of plant form and size, while using equipment available to most biology labs. Fig.1 illustrates the system framework. Briefly, the system can be divided into three main modules: SFM, stereo matching and depth optimization and merging. Unlike other SFM methods that consider input images as an unordered set, we organize the input images into a sequence of images. We directly select the starting pair of images to begin an incremental SFM to estimate the camera parameters as described in our previous work [11], and then rectify the stereo pairs by depth-map computation and optimization. Finally, all the refined depth-maps are merged together into a 3D point cloud to provide a final reconstruction. We elaborate on each of the steps as follows:

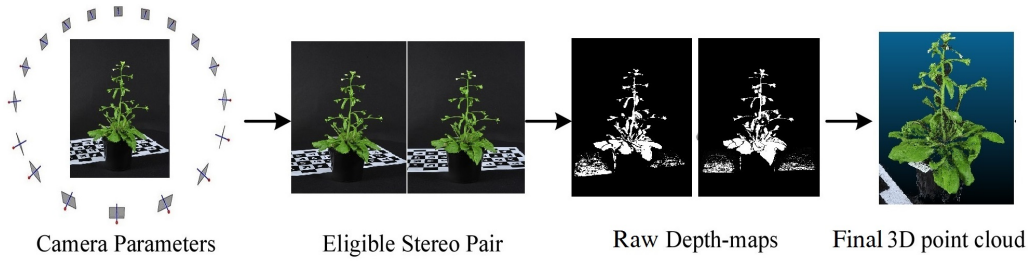

Fig. 1. 3D reconstruction framework

2.1 Structure from Motion

Given a short baseline image/video sequence I with n frames taken by a turntable or freely moving camera, we denote $I = \{I_t | t = 1, 2, ...n\}$, where I_t represents the color image captured at the frame t. The set of camera parameters for frame t in an image sequence is denoted as $C_t = \{K_t, R_t, T_t\}$, where K_t is the intrinsic matrix, R_t is the rotation matrix, and T_t is the translation vector. The parameters can be estimated reliably by the SFM techniques [5] [16]. Our system employs the SFM method of Snavely et al. [16] via the open source software *Bundler*. We improve the *Bundler* at two important steps: i) We sort the images as an ordered sequence according to incremental image name, and thereby reduce the computational cost from $O(n^2)$ to $O(n)$ in feature matching

procedure ; ii) We speed up the SIFT features detection based on GPU hardware, which is around 10 times faster than CPU. In addition, as alternative methods, we have also tried to use ASIFT(Affine-SIFT) [12] feature detection and matching in SFM. Although more features can be detected and matched, the ASIFT does not significantly improve the experimental results.

2.2 Stereo Pair Selection

Not all image pairs are eligible for stereo matching. The selection of stereo image pairs is important not only for the accuracy of the final MVS result but also for the time performance of the system. A good candidate neighboring image pairs should have sufficient ray intersection angles with the reference images, and have a suitable baseline neither too short to degrade the reconstruction accuracy nor too long to have less common coverage of the scene. Hence, the pair selection is based on two statistics from i-th possible image pair: the angle between principal view directions and the distance between camera optical centers.

To select eligible stereo pairs, we supposed to have n frames, and for the i-th one, we computed θ_{ij} ($j = 1, 2, ...n$) which is the angle between principal view directions of cameras i and j, and d_{ij} ($j = 1, 2, ...n$) which is the distance between optical centers of cameras i and j. In our system, eligible stereo pairs are determined by the following rules:

i) Firstly, θ_{ij} satisfies $3° < \theta_{ij} < 15°$.

ii) Secondly, we compute the median $\bar{d}$ of d_{ij}, and remove the stereo pair whose $d_{ij} > 2\bar{d}$ or $d_{ij} < 0.05\bar{d}$.

iii) Finally, each remaining image view i has at most two stereo pairs with its neighbours j according to the metric d_{ij}.

Suppose the number of images is n, the number of eligible stereo pair is less than $2n$.

2.3 Depth Maps Computation

Let I_b and I_m be a eligible image pair, our objective is to estimate a set of depth maps. The classic SGM algorithm [6] aims to recover disparities across stereo pairs by minimizing the following global cost function:

$$E(\mathbf{D}) = \sum_{\mathbf{x}_b}(C(\mathbf{x}_b, \mathbf{D}(\mathbf{x}_b)) + \sum_{\mathbf{x}_N} P_1 T[\|\mathbf{D}(\mathbf{x}_b) - \mathbf{D}(\mathbf{x}_N)\| = 1] + \sum_{\mathbf{x}_N} P_2 T[\|\mathbf{D}(\mathbf{x}_b) - \mathbf{D}(\mathbf{x}_N)\| > 1])$$

Where $\mathbf{D}(\mathbf{x}_b)$ represents the disparity estimations of all base image pixels $\mathbf{x}_b$ of I_b, $\mathbf{x}_N$ denotes base image pixels in the neighborhood of $\mathbf{x}_b$, P_1 and P_2 are penalty constants to control the gain of a little or a larger disparity changes respectively, T is an operator evaluating to one if the subsequent condition is true and evaluate to zero else. $C(\mathbf{x}_b, \mathbf{D}(\mathbf{x}_b))$ computes the pixel-wise similarity measures. Our method is based on the OpenCV API which implements a memory and time efficient modification of SGM, where a hierarchical approach was proposed to initialize and refine the Mutual Information (MI) matching cost, and initial disparity images were computed by matching high level (low resolution) image pyramids. The resulting disparities were then used to refine the MI matching cost for processing the subsequent pyramid level.

2.4 3D Reconstruction

Since the raw depth maps may not completely agree with each other due to depth errors, a refinement process is necessary to enforce consistency over neighboring views. For each point p in image I_t, we back projected it to 3D using its depth value and the camera parameters. Therefore erroneous disparities were filtered additionally by checking for geometric consistency in object space. After refinement, all the depth-maps could be integrated into a 3D point cloud to represent the object scene. While the depth-maps may contain lots of redundancies, because different depth-maps may have common coverage of the object scene in neighboring images, we do not consider removing the redundancies for the sake of cost-effectiveness at the moment.

3 Experimental Results

Plant images were mainly captured using Canon digital cameras (Canon 600D) with $18 - 55mm$ lens, and Nikon digital cameras (D60 and D700) with $28 - 105mm$ lens and network camera with 50mm lens. The images were stored in JPG format with 3184×2120 or 3696×2448 high-resolution. Video were captured at 1920×1024 with AVI format. In order to keep even illumination and uniform background, two diffuse lighting rigs and a black backdrop were used. The method was run on an Intel i7 laptop (Dell Precision M6700 with 16G ram and Nvidia graphics card).

Fig.2 shows the 3D reconstruction of the plants produced by the proposed method. Fig.3 shows the 3D reconstruction of an Arabidopsis from 66 images , produced by the existing state of the art methods [3], [8] and our method respectively. Generally, the running time of our method was 10-30 minutes depending on the number of images. In the case of Fig.3, Our method ran 15 minutes to produce the final 3D point cloud from 66 input images. The CMPMVS software from [8] was very slow and ran around 182 minutes. The PMVS software from [3] was faster than our method and only ran 3 minutes but quality was poor. The proposed method is able to produce more dense and complete 3D point cloud of plants than previously reported methods.

In order to evaluate the quality of 3D reconstruction, we have extracted the plant height, stem width, leaf area, branch angle from resulting 3D point cloud of plant in case study, and compared results with limited ground truth that came from the manual measurement or commercial 2D image measuring software. Fig.4 shows an Arabidopsis and the produced 3D point cloud, and Table.1 shows the comparison errors of some leaves and main stem.

Experimental results show that the accuracy of the proposed method is high enough to distinguish different genotypes plant and precision is stable enough to adapt slightly deviated experimental conditions.

With the proposed method, no extra camera calibration or other strict environmental conditions is required, and data acquisition times are in the range 1-2 minutes per specimen, depending on plant complexity. The camera parameters, such as focus length, depth of field and time of exposure, etc., can be adapted

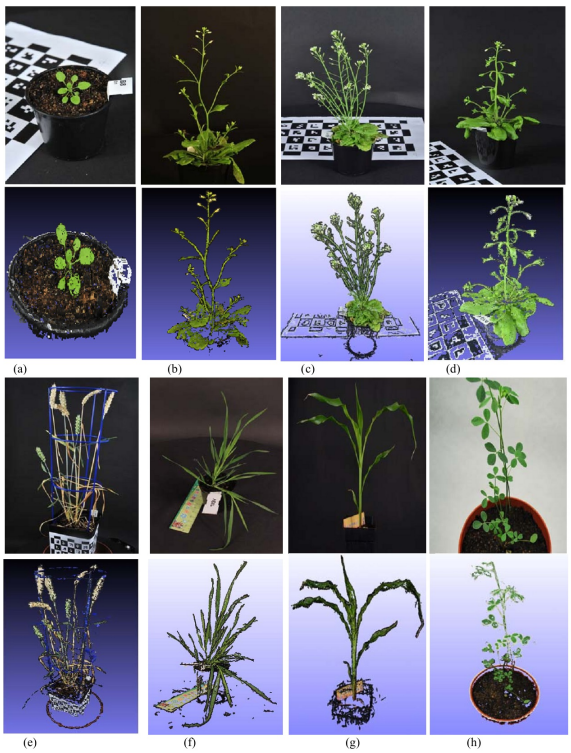

Fig. 2. Raw image (top) and 3D reconstruction (bottom) of diverse plants. (a)-(d) Arabidopsis strains and mutants. (e) wheat. (f) Brachypodium. (g) maize. (h) clover.

Fig. 3. 3D reconstruction of an Arabidopsis. (a) raw image (one of 66 images). (b)(c)(d) 3D results by CMPMVS [8], PMVS [3] and the proposed method respectively.

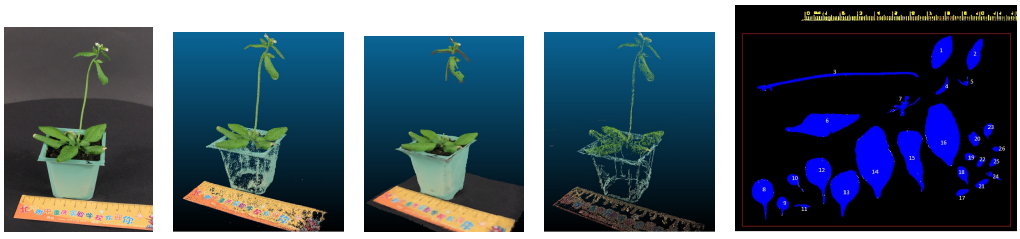

Fig. 4. (a) Raw image of an Arabidopsis (one of 62 images) (b)(c)(d) 3D reconstruction by the proposed method, CMPMVS [8], PMVS [3] respectively. (e) Measurement of the leaves and stem using WinDas scanner and software.

Table 1. Errors in different parameters of an Arabidopsis using different methods

Item	Ground truth (mm)	Proposed method (mm)	Ref [3] (mm)	Ref [8] (mm)
Length of leaf1	21.5	-1.2	N/A	N/A
Width of leaf1	10.7	0.9	N/A	N/A
Length of leaf2	19.5	-0.1	N/A	N/A
Width of leaf2	7.6	0.6	N/A	N/A
Length of leaf14	41.7	-0.5	N/A	2.2
Width of leaf14	21.1	1.0	N/A	3.5
Length of leaf16	36.7	-1.7	N/A	2.6
Width of leaf16	19.3	-0.9	N/A	3.1
Length of main stem	84.8	-2.1	-3.4	N/A
width of main stem	1.4	-0.1	NA	N/A
Root mean square error (RMSE)		1.1		
Mean absolute percentage error (MAPE)		4.7%		

widely to obtain high quality and clear images with an optimal resolution from a diverse set of plants. Only basic photography skills are required for effective capture and storage of data.

4 Conclusion

In this paper, we demonstrated that 3D reconstruction based multi-view images can be employed as a low-cost, powerful alternative for non-destructive plant phenotyping. The proposed dense 3D reconstruction method excelled in producing accurate 3D point cloud of various plants while retaining colors, textures, shapes as compared to other current methods [3][8]. It is flexible and efficient. Abundant useful phenotypic data as plant height, plant topology, stem width and length, numbers of leaves, leaf area and leaf angle, etc. can be extracted.

The main limitations of the proposed method are (i) the final 3D model of plants was only pure point clouds without meshes so that it cannot be directly processed in the common commercial CAD/3D software (ii) there are still some gaps or holes in the final 3D model of the plants due to occlusions, texture-less regions and blurred images. In addition, the proposed method is still relatively slow, especially when processing hundreds of images. Therefore, future work will aim at improving computational performance and accuracy.

Acknowledgments. We acknowledge funding from the European Union FP7 Capacities Programme (Grant Agreement No. 284443: European Plant Phenotyping Network, an Integrating Activity Research Infrastructure project) and BBSRC NCG Grant Ref: BB/J004464/1.

References

1. Biskup, B., Scharr, H., Rascher, U.S.: A stereo imaging system for measuring structural parameters of plant canopies. Plant, Cell & Environment **30** (2007)
2. Cobb, J.N., et al.: Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theoretical and Applied Genetics **126**, 867–887 (2013)
3. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. IEEE Tran. PAMI **32**, 1362–1376 (2010)
4. Granier, C., et al.: Phenopsis, an automated platform for reproducible phenotyping of plant responses to soil water defficit in arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deffcit. New Phytologist **169**, 623–635 (2006)
5. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2004)
6. Hirschmuller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Trans. PAMI **30**(2), 328–341 (2008)
7. Ivanov, N., et al.: Computer stereo plotting for 3-d reconstruction of a maize canopy. Agricultural and Forest Meteorology **75**, 85–102 (1995)
8. Jancosek, M., Pajdla, T.: Multi-view reconstruction preserving weakly-supported surfaces. In: Proc. CVPR, pp. 3121–3128 (2011)
9. Kaminuma, E., Heida, N., et al.: Automatic quantification of morphological traits via three-dimensional measurement of arabidopsis. Plant **38**, 358–365 (2004)
10. Li, Y.Y., Fan, X.C., Mitra, N.J., Chamovitz, D., Cohen-Or, D., Chen, B.Q.: Analyzing growing plants from 4d point cloud data. ACM Trans. Graphics **32** (2013)
11. Lu, L., Yonghuai, L., Minglan, S., Jiwan, H., John, H.D.: A cost-effective automatic 3D reconstruction pipeline for plants using multi-view images. In: Proc. 15th Towards Autonomous Robotic Systems (2014)
12. Morel, J.M., Yu, G.: Asift: A new framework for fully affine invariant image comparison. SIAM J. Img. Sci. **2**(2), 438–469 (2009)
13. Paproki, A., Sirault, X., Berry, S., Furbank, R., Fripp, J.: A novel mesh processing based technique for 3d plant analysis. BMC Plant Biology **12** (2012)
14. Quan, L., Tan, P., Zeng, G., Yuan, L., Wang, J.D., Kang, S.B.: Image-based plant modeling. ACM Trans. Graphics **25**, 599–604 (2006)
15. Reuzeau, C.: Traitmill (tm): A high throughput functional genomics platform for the phenotypic analysis of cereals. Vitro Cellular & Developmental Biology-Animal **43**, S4–S4 (2007)
16. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3d. In: Proc. SIGGRAPH, pp. 835–846 (2006)
17. Santos, T., Oliveira, A.: Image-based 3d digitizing for plant architecture analysis and phenotyping. In: Proc. Workshop on Industry Applications in SIB-GRAPI (2012)
18. Yamazaki, S., et al.: The theory and practice of coplanar shadowgram imaging for acquiring visual hulls of intricate objects. IJCV **81**, 259–280 (2009)