

Plant Phenotyping: An Active Vision Cell for Three-Dimensional Plant Shoot Reconstruction^{1[OPEN]}

Jonathon A. Gibbs,^{a,2} Michael Pound,^a Andrew P. French,^a Darren M. Wells,^b Erik Murchie,^b and Tony Pridmore^{a,3}

^aSchool of Computer Science, University of Nottingham, Jubilee Campus, Nottingham NG8 1BB, United Kingdom

^bSchool of Biosciences, University of Nottingham, Sutton Bonington Campus, Sutton Bonington, Leicestershire LE12 5RD, United Kingdom

ORCID IDs: 0000-0002-2772-2201 (J.A.G.); 0000-0002-5016-1078 (M.P.); 0000-0002-8313-2898 (A.P.F.); 0000-0002-4246-4909 (D.M.W.); 0000-0002-7465-845X (E.M.); 0000-0002-9485-1978 (T.P.)

Three-dimensional (3D) computer-generated models of plants are urgently needed to support both phenotyping and simulation-based studies such as photosynthesis modeling. However, the construction of accurate 3D plant models is challenging, as plants are complex objects with an intricate leaf structure, often consisting of thin and highly reflective surfaces that vary in shape and size, forming dense, complex, crowded scenes. We address these issues within an image-based method by taking an active vision approach, one that investigates the scene to intelligently capture images, to image acquisition. Rather than use the same camera positions for all plants, our technique is to acquire the images needed to reconstruct the target plant, tuning camera placement to match the plant's individual structure. Our method also combines volumetric- and surface-based reconstruction methods and determines the necessary images based on the analysis of voxel clusters. We describe a fully automatic plant modeling/phenotyping cell (or module) comprising a six-axis robot and a high-precision turntable. By using a standard color camera, we overcome the difficulties associated with laser-based plant reconstruction methods. The 3D models produced are compared with those obtained from fixed cameras and evaluated by comparison with data obtained by x-ray microcomputed tomography across different plant structures. Our results show that our method is successful in improving the accuracy and quality of data obtained from a variety of plant types.

With the population increasing and expected to reach 9 billion within the next four decades, it is no wonder that demand for food is increasing (Sticklen, 2007; Faaij, 2008; Paproki et al., 2012). Moreover, developing countries, such as China and India, are increasing food intake per capita and driving the demand for a richer, more varied diet, such as meats and dairy. Climate change, leading to more frequent and severe flooding, and a shortage of arable land constitute additional challenges. Furthermore, it has been predicted that, without crop climate adaptation, the production of food will deteriorate (Adeloye, 2010; Challinor et al.,

2014). In order to deal with such demands, innovative approaches to increasing agricultural production are necessary.

Connections between the underlying genetic code and the visible physical structures and functions of plants (i.e. phenotyping) can aid in the identification of more productive crop species. A comprehensive understanding of plant phenotypes informs breeding and genetic selection, facilitating, for example, more effective nutrient use and photosynthetic activity, thereby increasing crop yield and stability across more extreme environments (Quan et al., 2006). The relationship between phenotype and genotype has received an increased amount of attention in recent years, with significant progress made in the study of genetics. The recovery and analysis of traits such as plant growth, development, and tolerance, however, remains a serious bottleneck (Furbank and Tester, 2011). Two-dimensional (2D) approaches to plant phenotyping have been used extensively, although they have numerous limitations, most notably the inability to accurately reflect 3D quantities. For example, a curved leaf in a 2D image will have a significantly smaller surface area than in a 3D model. 2D methods struggle to capture plant structure, and accurate measurement of growth is challenging. The use of 3D models overcomes many of these difficulties, allowing more and more traits to be obtained accurately. Once a 3D model of a given plant has been built, it can be reanalyzed, should new trait measurements be required. This may

¹This work was funded by Engineering and Physical Sciences Research Council PhD Studentship Award 1499261 (to J.A.G.) and Biotechnology and Biological Sciences Research Council Grant BB/R004633/1, "The 4-Dimensional Plant: Enhanced Mechanical Canopy Excitation for Improved Crop Performance."

²Author for contact: psxjg6@nottingham.ac.uk.

³Senior author.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Jonathon A. Gibbs (psxjg6@nottingham.ac.uk).

J.A.G. designed and implemented the system, performed the experiments, and wrote the article; M.P. supported with 3D reconstructions; A.P.F., D.M.W., and E.M. supported with biological relevance and fit for purpose; T.P. supervised the project.

[OPEN]Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.18.00664



Figure 1. The AVC, composed a Canon 650D camera, a UR5 universal robot, and an LT360EX turntable upon which the plant is placed.

not be possible in 2D approaches, where image acquisition often is designed to provide a particular, limited set of data. Access to accurate 3D models also supports simulation-based studies of plant functions, such as photosynthesis (Burgess et al., 2015, 2017).

The construction of accurate 3D models of plants is extremely challenging. Existing approaches fall into the two categories of rule based or image based (Remondino and El-Hakim, 2006). Rule-based approaches use knowledge of plant structure, forming and generating example models consistent with that knowledge. Although rule-based approaches can produce satisfactory results, their use often requires expert knowledge, and rules usually are targeted toward specific plant types. Plant structure also varies significantly across species and environments, making it difficult to predict structures *a priori*. More importantly, although they can generate visually realistic models, the representations produced may not correspond to any real, existing plant. Consequently, rule-based models are unsuitable for high-resolution phenotyping tasks. In contrast, image-based methods develop accurate 3D models of real, viewed plants. These models can

be used to support both simulations of plant function and the extraction of trait measurements (Burgess et al., 2015, 2017).

One of the more popular approaches to 3D modeling is multi-view stereo (MVS). Here, a number of images (several tens) are captured from distinct viewpoints. Given sufficient overlap between views, it is possible to match features between images and produce a 3D point cloud, to which a surface can be fitted. Although MVS has been successful in a variety of domains, plants are particularly challenging objects to model. Individual leaves can be very similar in appearance and densely packed, occluding each other from many viewpoints. They often lack the surface texture needed when matching image features, assuming local coherence and smoothness. The leaves of many species also are highly reflective, making alternative laser scanning approaches less effective. For a review of 3D modeling algorithms for plants, readers are encouraged to see Gibbs et al. (2017).

The high-throughput phenotyping systems deployed in plant and crop science are now routinely gathering large numbers of images from which 3D models might

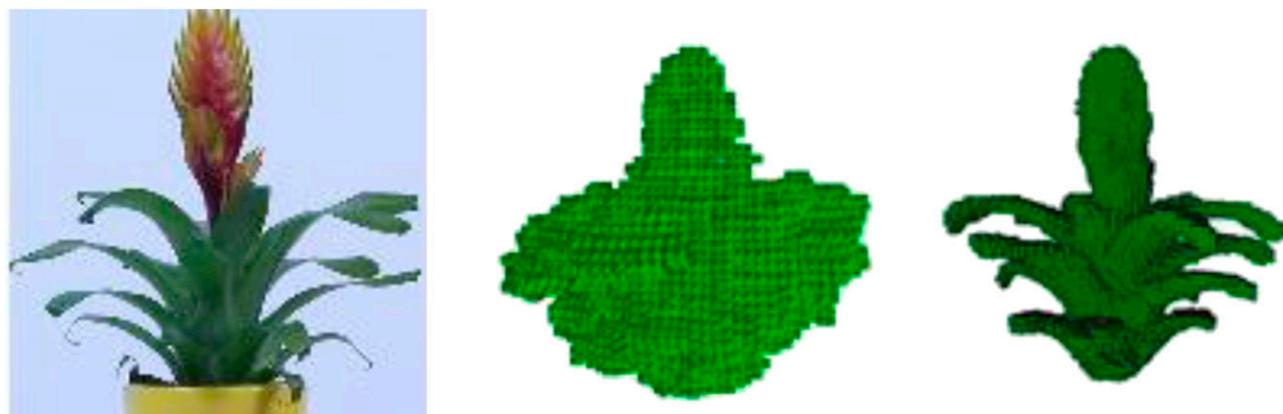


Figure 2. Initial representation. Left, an original image of a target plant (bromeliad); middle, the initial representation after 10 images; right, the final voxel model showing more object features after acquiring additional viewpoints.

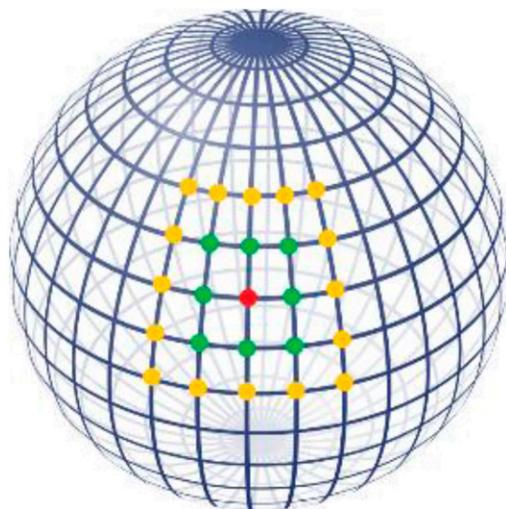


Figure 3. The view sphere representation, which encloses the plant being modeled such that it is centered. The red dot is an example of an initial optimal viewpoint; should this fail, it is expanded to green, then to yellow, and so on.

be obtained. Current installations, however, typically rely on fixed viewpoints that are not adapted to the specific plant being examined or are designed with one species in mind. Some systems rotate the plant during imaging but still use static camera positions. The relation between viewpoints and plant, therefore, remains fixed, regardless of the structure of the plant, which may vary widely. This means that, in many cases, the images captured are far from optimal for the given plant. In order to capture 3D models useful for phenotyping, there is a need for a more intelligent image-capture system optimized for 3D reconstruction and sensitive to variations in plant architecture.

In this work, we show that active computer vision (Aloimonos et al., 1988) can aid the reconstruction of complex plants by providing reactive, and therefore improved, image-acquisition strategies. Active vision systems automatically control and manipulate camera viewpoints to gather information to best support the task at hand. Active vision methods already have played a role in other plant-related tasks. For example, Hemming et al. (2014b) attached a camera to a robot arm in order to identify peppers (*Capsicum annuum*) to be collected. The effect of camera placement on fruit picking also has been investigated (Hemming et al., 2014a), with active vision used to address the problem of occlusion. However, the process of capturing images for 3D reconstruction, known as image selection, currently is an insufficiently considered resource in image-based 3D reconstruction (Hornung et al., 2008).

We propose a framework to automatically capture a set of images suitable for use in 3D modeling, via MVS, of different and contrasting plant structures. This work directly addresses the competing demands placed on

image acquisition: too many images can introduce redundancy and result in excessive processing times, while too few images result in an incomplete model. We identify a set of viewpoints that enable a reliable 3D model to be reconstructed without scanning the plant excessively. We present a solution suitable for deployment in an automated, high-throughput phenotyping system. This article describes a fully automated, active vision cell (AVC) that is capable of manipulating a camera's viewpoint to produce high-quality 3D models of a wide range of plants by adapting to the visual information available, without user intervention. The approach described here offers more flexibility than existing large-scale phenotyping systems by adapting to the natural variation of individual plants. This is achieved by investigating an initial, crude representation of plant structure in order to reposition the camera and obtain improved data.

SETUP/METHOD DEVELOPMENT

The accuracy and reliability of 3D models depend heavily on the quality of images, while its computational requirements are dependent on the number of images. Images do not contribute equally to the quality of a reconstruction: some are redundant, while others add large amounts of high-quality, necessary data (Seitz et al., 2006). Here, we propose an AVC designed to provide sufficient data to ensure a reliable representation without the need for specific expertise on the part of the user, with the ability to adapt to different

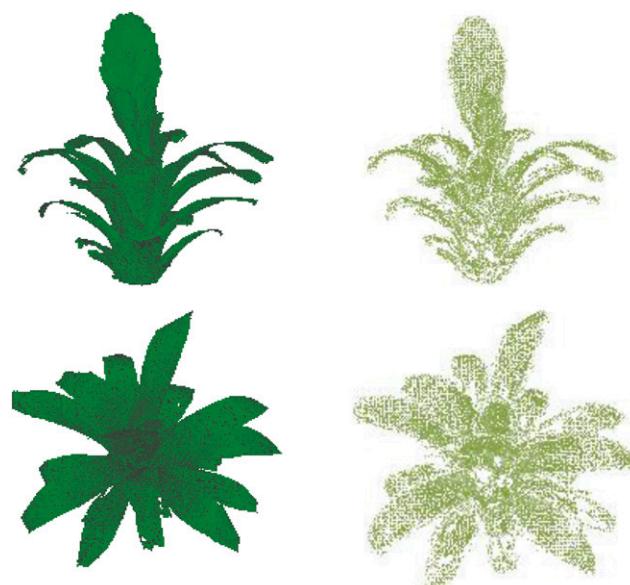


Figure 4. Experiment 1 conducted on a bromeliad. The first column is the x-ray data, obtained using a CT scanner. The top row presents a side view and the bottom row presents a top-down view. The second column is a point set obtained using the AVC proposed here.

Table 1. Experiment 1 results, bromeliad

Method	Mean	SD	Points	Images	Per Image
One static	0.3574	1.1795	141,073	40	3,526.80
Two static	0.3442	0.7450	155,396	40	3,884.90
Arbitrary	0.2693	1.1267	227,338	40	5,683.50
AVC	0.1959	0.6773	290,236	36	7,637.80

plant structures, and without analyzing excess numbers of images.

Cell Design and Calibration

Our AVC is composed of three main components: a high-precision turntable (LT360EX; Linear X Systems) with a resolution of 0.0015° , a robot arm providing 6° of freedom (UR5; Universal Robots), and a standard color camera (Canon 650D; Canon) mounted on the robot arm (Fig. 1). A single software interface is used to control each of the hardware components. The UR5 is sent commands using strings via sockets, the LT360EX is controlled using serial communications, and the Canon 650D is controlled via a Software Development Kit.

Calibration, the process of obtaining reliable 3D camera parameters for each view, is an important first step in any 3D reconstruction pipeline. Calibration usually is an automatic process, determining the physical parameters of each hardware component and quantifying the relationships between them and the viewed environment. The calibration process can be organized into four stages: camera calibration, robot calibration, calibration of the remaining unknowns, and turntable calibration. All four calibration steps are required to determine the position of the camera for active vision. In simple terms, the calibration aims to estimate the position and orientation of each component in the setup (the robot and turntable) and the camera lens and sensor.

Camera Calibration

Camera calibration is used to estimate the intrinsic and extrinsic parameters of the camera, which are used to determine its location for the calibration of the robot. A standard checkerboard calibration target, in which the dimensions of the squares are known, is placed on the turntable. Given a series of images of this calibration object at distinct viewpoints, it is possible to recover the position, orientation, and internal parameters of the camera that captured each image. Internal parameters often are termed intrinsic parameters and consist of the focal length, offset, and axis skew. The 3D plant models produced are expressed in world coordinates with respect to a coordinate frame located on the checkerboard. The bottom right corner of the checkerboard is the world origin $(0, 0, 0)$. Camera calibration provides a transformation between world coordinates and a coordinate frame centered on the camera. This

transformation can be used to project any 3D world position into a 2D camera position in its image frame.

Robot Calibration

Robot calibration estimates the position and orientation of the end of the robot arm (i.e. the end effector). Also known as forward kinematics, robot calibration is achieved using a simultaneous closed-form quaternion approach (Dornaika and Horaud, 1998). This produces a transformation matrix specifying the relationship between the base of the robot and the end effector. This transformation matrix provides the rotation and translation needed to transform one robot position to another.

Calibration of Unknowns

After transformations linking the base of the robot to the camera and the camera to the world (turntable) are available, it is possible to calculate the relationship between the base of the robot and the turntable (world). The remaining calibrations can be calculated as a linear equation in the form $AX = YB$, where A (the world to camera) and B (the robot base to the end effector) are now known and where Y (the world to robot base) and X (the camera to the end effector) are the two unknowns. A closed-form approach to the linear equation

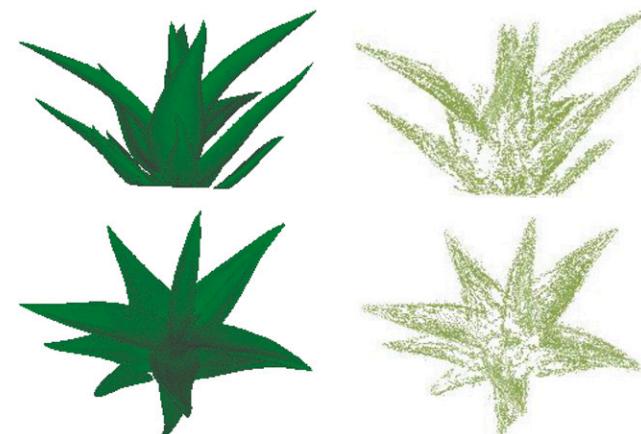


Figure 5. Experiment 2 conducted on aloe. The first column is the x-ray data, obtained using a CT scanner. The top row presents a side view and the bottom row presents a top-down view. The second column is a point set obtained using the AVC proposed here.

Table 2. Experiment 2 results, aloe

Method	Mean	SD	Points	Images	Per Image
One static	1.4517	3.6624	159,870	40	3,996.80
Two static	1.6911	3.6143	160,592	40	4,014.80
Arbitrary	1.8963	4.5674	183,027	40	4,575.70
AVC	1.3329	3.5930	216,791	31	5,705.00

has been used to determine the remaining unknowns (Dornaika and Horaud, 1998).

Turntable Calibration

Rotating the turntable, which is necessary to provide complete access to the plant, changes the relationship between robot/camera and world coordinates. To calibrate the turntable, it is rotated by 90° four times. The camera is recalibrated each time, giving four positions for the world coordinate origin. Plotting the four origins obtained from the calibration in two dimensions and connecting the diagonal origins using a straight line allows the center of rotation to be solved as a line intersection problem. The center of rotation is used to calculate a new world coordinate frame each time the turntable is rotated. At this point, we have a fully parameterized relationship between the camera system, robotic arm, and turntable.

Active Image Acquisition

There are two stages to 3D modeling within the AVC: the first requires the creation of a crude, initial plant model, represented by a series of voxels; the second stage involves an analysis of this initial representation to identify undersampled and oversampled (imaged) regions of the plant. The robot arm then is directed automatically to acquire more data, while unnecessary images are removed. Note that the images used to construct the volumetric proxy also are determined automatically, on the basis of 2D image features, as described below.

An Initial Volumetric Plant Representation

To acquire an initial volumetric representation of a plant, we capture a series of images. These are taken from automatically determined camera locations circling the plant at three different heights. The first image is acquired after positioning the camera so that its principal axis (line of sight) lies in the plane of the turntable and passes through its center of rotation. A Euclidean color filter, which filters pixels where the color is inside or outside of a red, green, and blue sphere with a specified center and radius, is applied to separate plant pixels from the white background. We then apply three simple rules to move the camera to center the plant (which may be of arbitrary size, asymmetric, etc.) within the camera's field of view (FOV): (1) if there is too much white space surrounding the plant

region (i.e. if the distance from the plant region to the edge of the image is greater than a specified threshold), move the camera forward; (2) if one side of the plant is outside the camera's FOV, move the camera laterally to ensure it is inside; and (3) if more than one side is outside the camera's FOV, move the camera backward. The resulting camera location forms the starting point for image acquisition. Once an acceptable viewpoint has been determined a series of images is captured by rotating the plant and acquiring an image every 36°, producing 10 images with the camera fixed at the initial elevation.

Space carving (Seitz, 2000) is used to generate the initial 3D model from the first image sequence. Space carving operates by projecting the silhouette of the target object (the plant) into 3D space to define the volume possibly occupied by the object. Projecting silhouettes extracted from multiple images, and taking the intersection of the volumes they produce, reduce the size of this volume, creating an increasingly more accurate model.

This 10-image model of a complex plant (Fig. 2) is of limited value, but it does allow an estimate of the plant height to be made. The camera is raised to be level with the top of the plant, automatically recentered as described above, and another 10 images are acquired by rotating the turntable. This is known as the level

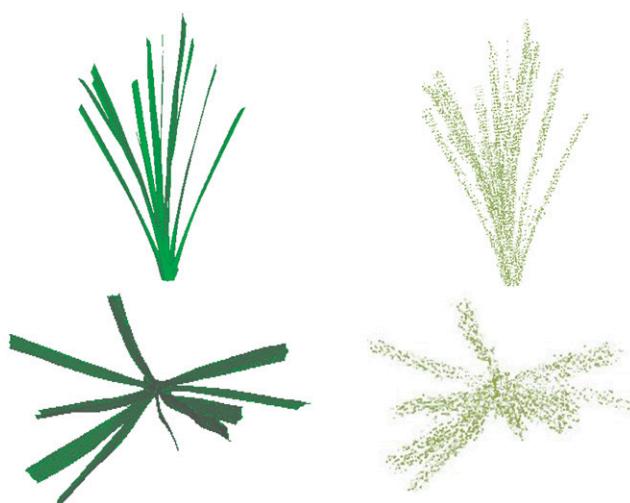


Figure 6. Experiment 3 conducted on a cordyline. The first column is the x-ray data, obtained using a CT scanner. The top row presents a side view and the bottom row presents a top-down view. The second column is a point set obtained using the AVC proposed here.

Table 3. Experiment 3 results, cordyline

Method	Mean	SD	Points	Images	Per Image
One static	0.7565	2.0167	122,851	40	3,071.30
Two static	0.8638	2.8122	94,193	40	2,354.80
Arbitrary	1.0284	4.7614	80,154	40	2,003.90
AVC	0.7384	2.0691	143,049	26	3,764.40

2 position, having moved up along the z axis in one increment, where the first set of images were captured at level 1, in line with the turntable. To improve coverage, the turntable is rotated 12° before image acquisition begins. This means that the level 1 and 2 camera positions are not aligned vertically but offset by 12° . The new images then are used to refine the volumetric model and, therefore, plant height estimation.

To complete the volumetric representation, the camera is raised to twice the newly estimated height of the plant, a further 12° offset is added, and a final 10 images are acquired. By increasing the height of the camera to above the height of the plant, it is possible to get a set of top-down images uncovering new information, particularly useful for plants with wide flat leaves, such as broadleaf species including legumes and squashes.

This image-acquisition strategy is designed to achieve a set of varying viewpoints that sample the area around the plant while keeping the plant in view. Note that we do not recenter the plant in each image, only in the first image captured at each level. However, given plants with a high degree of asymmetry, the rules above could be applied after each rotation of the turntable.

The final volumetric model remains comparatively crude and low resolution, giving a blocky appearance, and is unable to represent some features at all, such as concavities. However, it does provide a sufficient intermediate representation for evaluation via forward ray tracing (Vasquez-Gomez et al., 2013), in which rays from the camera are projected into the scene to determine the intersection with the object, and so determines which cameras can see which parts of the developing 3D model.

Plant Model Refinement

The next step is the automatic refinement of the image set, removing those that are unnecessary and obtaining further images of underrepresented sections of the plant. Images are removed if each voxel in the plant proxy representation is still seen by more than three cameras after their removal. In practice, MVS produces higher quality results when an area has been seen three times or more.

View planning is then performed to determine which additional data to capture. Traditionally, view planning evaluates each possible view on a per voxel basis: each voxel is evaluated independently for every possible camera position in the view sphere (Massios and Fisher, 1998; Wong et al., 1999). If we were to do this in our cell, and if we limit robot movements in whole

degrees, it is possible to move 180 points from top to bottom and 360 points around the view sphere, resulting in 64,800 camera positions that would require evaluation. We reduce this complexity by clustering voxels together and evaluating specific views on a per-cluster basis. There are four stages here: (1) clustering, (2) cluster evaluation, (3) camera placement, and (4) data acquisition.

1. *Clustering*. Each voxel is represented by a single point lying at its center, and the k nearest neighbor (k -NN) algorithm is used to cluster the point set. k -NN is a simple machine-learning algorithm that clusters the point set into a series of k nearest neighbors. That is, points are added to some cluster that are within the range of the centroid when given some radius. k -NN finds the k nearest neighbors to a point that are within some radius of the center of the cluster, the starting point. We implement this algorithm using a KD-tree data structure, which significantly improves performance when applying nearest-neighbor searchers to points in k dimensions.

2. *Cluster Evaluation*. Each cluster must be evaluated to determine whether additional images need to be captured and, thus, to ensure that the object is sufficiently scanned. We propose a simple evaluation method that operates on the number of views in which a cluster is visible and the angle between the cameras that have seen the cluster (Furukawa and Ponce, 2010). If a cluster has a low score, then we mark the cluster as requiring

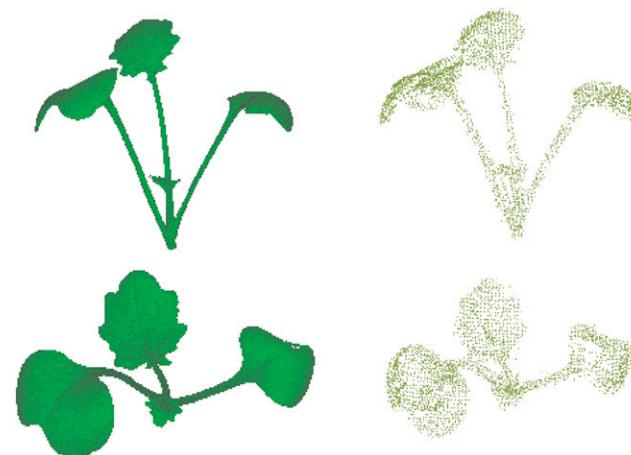


Figure 7. Experiment 4 conducted on brassica. The first column is the x-ray data, obtained using a CT scanner. The top row presents a side view and the bottom row presents a top-down view. The second column is a point set obtained using the AVC proposed here.

Table 4. Experiment 4 results, brassica

Method	Mean	SD	Points	Images	Per Image
One static	0.2007	0.7208	97,191	40	2,429.8
Two static	0.0867	0.4427	146,743	40	3,668.6
Arbitrary	0.1682	0.5466	178,418	40	4,460.5
AVC	0.0354	0.3912	349,311	21	16,633.9

additional viewpoints. The evaluation metric used is given in Equation 1:

$$Score = \frac{1}{C_n} \sum_{j=1}^{C_n} \left(\frac{seen(C_j)}{imgCrit} + \frac{maxAngle(C_{cam}, C_{cam})}{angleCrit} \right) \times 0.5 \quad (1)$$

where $seen(C_j)$ refers to the number of times each voxel has been seen in the cluster and $imgCrit$ is the number of times a point must be seen to ensure an accurate representation (we use 3 to match our Patch Based Multi-View Stereo (PMVS) settings). $maxAngle(C_{cam}, C_{cam})$ is the maximum angle between any of the cameras that can see the voxel, and $angleCrit$ is the minimal angle difference between cameras, to ensure different views (we use 20° , determined empirically).

We determine whether a cluster has been seen by a given camera via ray tracing. This simulates the projection of a ray of light from the camera to the cluster centroid. In order to improve performance, we implement a hierarchical ray tracing (HRT; Vasquez-Gomez et al., 2013) approach rather than a uniform ray tracing method. Uniform ray tracing traces dense rays through the scene irrespective of whether an intersection with a voxel occurs. HRT traces sparse rays, only increasing the resolution when voxels are touched by a ray. Starting at a coarse resolution, HRT continues until the maximum resolution is reached.

3. Camera Placement. Given a series of undersampled clusters, we proceed to calculate a series of viewpoints that can be used to capture additional information. We first determine the distance the camera is required to be from the object to ensure that the plant is completely within the FOV, without excess white space, using the camera parameters and object size. The size of our view sphere (Fig. 3) then is determined by Equation 2:

$$FOV = 2 \cdot atan\left(\frac{1}{2} \cdot \frac{s}{f}\right)$$

$$Distance = \frac{1}{2} \cdot \frac{\max(w, h)}{\sin(FOV)} \quad (2)$$

where s is the sensor size and f is the focal length, both of which are obtainable from the camera

specification. $\max(w, h)$ returns the maximum value of the object with respect to the height, h , and width, w .

Traditional view-planning methods evaluate every possible position on the view sphere; we significantly reduce the heavy computational requirements this brings by incrementally expanding our search should a view fail. A starting camera position is defined as the intersection of the normal of the cluster with the view sphere. The view is evaluated for correctness in two ways: the first is to perform inverse kinematics to ensure that the robot is able to reach the position, the second is ray tracing from the camera position into the scene to ensure that the cluster is not occluded from this viewpoint. If either of the evaluations fails, we incrementally expand over the view sphere, first evaluating positions in green (Fig. 3) and then yellow, and so on, expanding outward from the starting position until an acceptable viewpoint is found. This process is performed for each cluster that requires additional viewpoints to be captured, until views of all clusters have been obtained.

4. Data Acquisition. Once we have a series of camera positions, additional images are captured as necessary, and PMVS (Furukawa and Ponce, 2010) is used to generate a point cloud that can support surface reconstruction.

EVALUATION AND DISCUSSION

Active Cell Evaluation

Having a more accurate set of points that closely represent the surface of some unknown object significantly improves the quality of any subsequent 3D model, as it more faithfully represents the actual shape of the object. Moreover, a larger number of points further facilitates the faithful reconstruction by providing more detail of the plant structure.

Table 5. Experiment 5 results, chilli

Method	Mean	SD	Points	Images	Per Image
One static	0.2380	0.9420	113,284	40	2,832.1
Two static	0.1843	0.4245	247,442	40	6,186.1
Arbitrary	0.2536	2.2226	199,023	40	4,975.6
AVC	0.1022	0.4584	285,381	28	10,192.2

Table 6. Experiment 6 results, pumpkin

Method	Mean	SD	Points	Images	Per Image
One static	1.1220	1.8674	715,222	40	17,880.6
Two static	1.2104	3.4723	517,039	40	12,926.0
Arbitrary	0.6982	1.8200	852,426	40	21,310.7
AVC	0.3588	1.3823	1,048,576	30	34,952.5

Ground Truth Model

In order to evaluate our AVC's point clouds, x-ray images of our target plants were obtained using a GE v|tome|x M scanner housed in the University of Nottingham's Hounsfield Facility. The v|tome|x M provides volumetric images with a voxel resolution of 5 to 150 μm and, more importantly, is not subject to the occlusion problems faced by visible light imaging. Although some x-ray segmentation tasks are highly challenging, plant material and air are easily separated in the density data provided by microcomputed tomography (µCT), and, following noise reduction with a median filter, plant material was identified by applying a user-defined threshold. A complete image of the plant is formed. The surface of each plant is then represented in a standard triangular mesh format, providing a data structure (i.e. a ground truth model) against which point clouds obtained from the AVC can be compared.

It is worth noting that, while the µCT scanner produces accurate, highly detailed models, it is ill suited for general use in phenotyping shoots due to size

restrictions, time requirements (typically taking hours to scan a single object, in comparison with minutes taken by the method here), and the exceptionally high startup costs. Moreover, thin structural areas of the plant still can be missed, resulting in an incomplete reconstruction. However, it is useful for creating 3D ground truth models with which to compare a visual imaging system, as occlusion is not a problem for x-ray µCT.

Comparative Image-Based Models

The AVC-derived model was compared with traditional static and arbitrary camera placements. Static setups use one or more cameras that remain fixed in place, irrespective of the plant being modeled. Typically, the plant is rotated and images are captured. In the experiments conducted in this work, the method one static refers to the use of a single static camera placed horizontally alongside the plant, such that the whole plant is visible in the camera's FOV. Two static uses two fixed cameras, using the same placement as one static and adding another camera placed higher, vertically, above the other such that a top-down view of the plant is obtained. Arbitrary refers to the process of

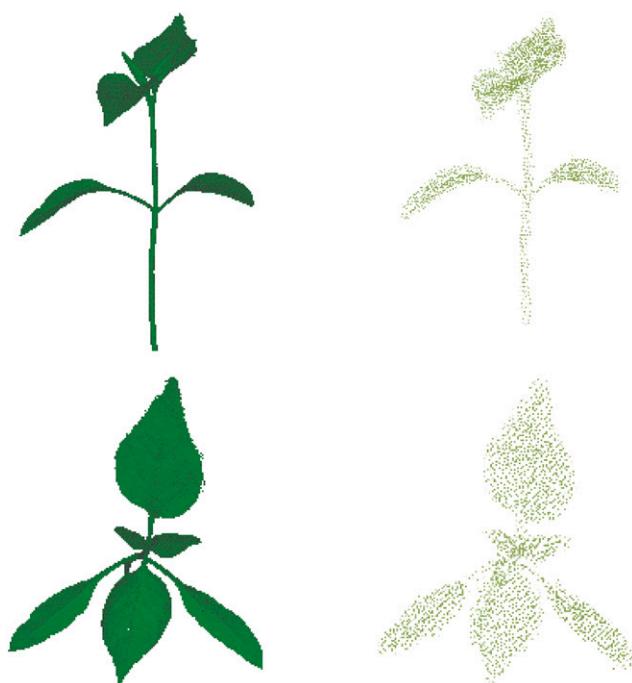


Figure 8. Experiment 5 conducted on a chilli plant. The first column is the x-ray data, obtained using a CT scanner. The top row presents a side view and the bottom row presents a top-down view. The second column is a point set obtained using the AVC proposed here.

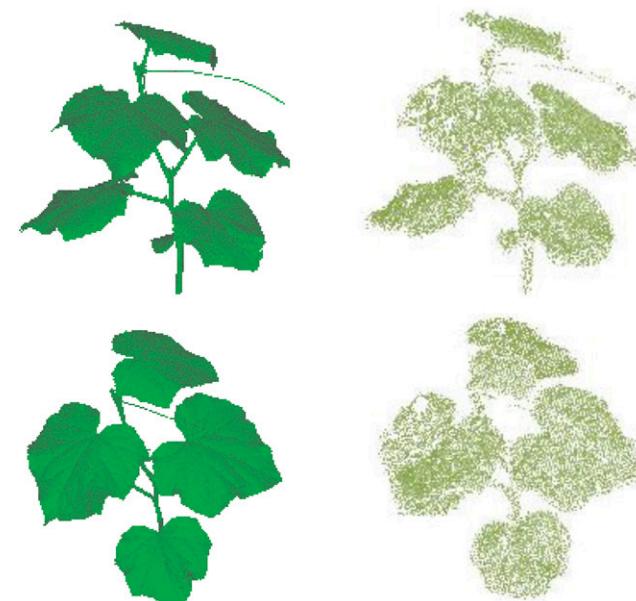


Figure 9. Experiment 6 conducted on pumpkin. The first column is the x-ray data, obtained using a CT scanner. The top row presents a side view and the bottom row presents a top-down view. The second column is a point set obtained using the AVC proposed here.

capturing images of the plant at distinct random positions and is commonly the method used when users manually capture images of plants.

Two evaluation metrics were employed: the number of points obtained and the distance from those points to the surface of the x-ray μ CT ground truth. Euclidean distance was used to determine the error of a point in the gathered data with respect to the surface of the ground truth. Six experiments were performed on plants varying in size, structure, and complexity: bromeliad (*Vriesea* sp.), aloe (*Aloe vera*), cordyline (*Cordyline* sp.), brassica (*Brassica napus*), chilli (*Capsicum* sp.), and pumpkin (*Cucurbita pepo*). The method is not limited to these plants and can be applied to plants that are much larger, such as wheat (*Triticum* sp.), maize (*Zea mays*), and barley (*Hordeum vulgare*), or other important crop species, with the only size restrictions relating to the reach of the robot arm.

Experiment 1

Experiment 1 was conducted on a bromeliad (Fig. 4). The Bromeliaceae are a family of monocot flowering plants in which over 3,400 species are known, native to the tropical Americas. While foliage takes different shapes and forms, the one used in this experiment is thin, broad, and flat. Consequently, views from above the plant, clearly seeing the wide leaves, will offer a great amount of insight into the plant size and structure. Occlusion, however, makes this problematic for static cameras that may be unable to see underlying leaf surfaces.

Table 1 compares the AVC approach with a static camera configuration. Mean refers to the distance of the points relative to the ground truth model; SD refers to the error of that distance; points refers to the number of points representing the 3D model and the number of points generated per image captured. When using a point cloud to drive a surface reconstruction approach (Pound et al., 2014), higher numbers of points allow a finer granularity on reconstructed surface patches, and a higher number of points per image indicates that more data can be generated for each image captured. Lower mean and SD also impact the quality of the surface reconstruction, where lower values illustrate a more accurate representation when compared with the ground truth. For the bromeliad, the AVC cell proposed here significantly outperforms the two static methods, obtaining more than 115% of the points in the first case, primarily due to the structure of the leaves, making it challenging for static cameras to view the leaf surface. In comparison with the arbitrary viewpoints, we see that we can increase the points per image by almost 35%, showing that intelligently selecting viewpoints in AVC improves performance despite fewer images; that is, we are obtaining more data per image. Furthermore, the reduction in the mean value by 27% shows that a more accurate point cloud is being produced (Supplemental Fig. S1).

Experiment 2

Experiment 2 was conducted on aloe (Fig. 5). The upward leaves occlude plant structure that lie directly behind them, making it challenging for views that are side on. Like the bromeliad from experiment 1, it consists of flat wide surfaces with little texture. Table 2 illustrates the results of the four image-acquisition methods.

From Table 2, we see that our AVC outperforms each of the standard methods, obtaining at least 18% more points while using 22.5% fewer images. The one static view obtains the fewest points, unable to deal with the concavities caused by the wide upright leaves. Two static also has fewer points; despite having two views, it is unable to obtain the data occluded by the outer leaves. Arbitrary viewpoints do overcome some of the occlusions but do not capture enough to deal with it completely. The AVC deals with the occlusions and recovers more accurate points with a reduced image set.

Experiment 3

Experiment 3 uses a cordyline, from a genus of approximately 15 species of monocotyledonous flowering plants in the family Asparagaceae (Fig. 6). Unlike the previous two experiments, experiment 3 focuses on a thin upright plant that is particularly crowded and occluded toward the base but relatively sparse toward the tips of the stems.

From Table 3, we see that our AVC significantly outperforms the arbitrary and two static views, but unlike the previous experiments, it has a smaller improvement over the traditional one static view. This highlights the fact that randomly adding images does not necessarily lead to an improvement and, in some cases, additional noise is added. As the plant contains few occlusions and has very thin nondrooping leaves, it is possible to capture a significant amount of information from a side view. However, despite the similarity of results between one static and our AVC points, our AVC uses 35% fewer images (26 relative to 40) than the single camera and obtains, on average, 22% more data per image used. This again shows that manipulating the viewpoint can improve accessibility to data and, thus, optimizes the processing power and time required to create a 3D model.

Experiment 4

Experiment 4 was conducted on brassica, an agriculturally important member of the Brassicaceae family (Fig. 7). This is a very small plant and, to avoid missing plant data, views need to be taken much closer than in the previous experiments. A traditional static image-acquisition strategy may struggle if not designed specifically for small plant species, as the camera will be positioned much farther away from the plant than necessary.

Table 4 indicates that the AVC captures more data despite using only half the images. This confirms that images in MVS reconstruction do not contribute evenly to the success of a reconstruction; rather, it is the quality of the images that has the greatest effect on the results.

Experiment 5

Experiment 5 was conducted using a chilli, which are grown widely in many countries as a cash crop (Fig. 8). Similar to experiment 4, the plant used was at an early developmental stage and, thus, of small size. Static cameras may miss data, particularly as the leaves and stems are thin.

Table 5 indicates again that the AVC is capable of capturing more, and, importantly, more accurate, data points from fewer images when compared with traditional methods. Although the two static camera approach does have a lower SD, it achieves this with many additional images.

Experiment 6

Experiment 6 was conducted using pumpkin (Fig. 9). The large flat leaves make occlusions for data acquisition a major problem, with the leaves often blocking the stem. Moreover, flat surfaces of plants often are problematic to reconstruct due to a lack of texture. Table 6 shows the results of the four approaches to image acquisition.

The large surface area results in the high number of points produced for this model (Table 6). As a result of the large surface area, with minimal texture, the SD for all methods is greater than that for previous experiments (see above). This is due to the difficulties associated with feature matching in PMVS. Despite this, the AVC still is able to produce an improved set of images with a smaller mean and larger set of points per image than any of the other methods.

Biological Application of the AVC Approach

Methods for the accurate 3D representation of plants (that also are accessible to many research groups) are increasingly important to basic and applied research for making new discoveries about plant function in addition to providing new traits for crop improvement. We still do not have a full understanding of how molecular and leaf-level events are scaled to the whole plant and field level and how this limits productivity. For example, there is a disconnect between phenotypes in growth rooms and those in more challenging field environments (Poorter et al., 2016). Nor is there a complete understanding of the canopy factors that cause variation in radiation use efficiency (Reynolds et al., 2000). The display of leaves to the sun and the way in which they influence the level of saturation of photosynthesis at each level is of huge importance to crop yield and optimizing architecture (e.g. by combining

leaf angle traits with leaf density and possibly movement; Long et al., 2006; Burgess et al., 2015, 2016, 2017). Rapid and accurate means to achieve high-resolution 3D reconstructions, such as the AVC described here, combined with more accurate ray tracing and physiological models will enable us to do that.

The approach described here requires minimal user input and can be applied to any plant type or structure, with the only limitation on size being the reach of the robot arm. It is more accurate and requires fewer images than previous, static imaging approaches (Tables 1–6) and offers more flexibility than existing large-scale phenotyping systems by adapting to the natural variation of individual plants. The method is automatic, with user input limited to changing the plant, and is relatively quick in image capture and analysis relative to other methods, taking minutes as opposed to hours. Moreover, the method has reduced setup and running costs compared with some phenotyping systems, such as x-ray μ CT scanning.

CONCLUSION

We proposed an AVC for automatically capturing color images of plants in a controlled environment, with a view to using them for 3D model reconstruction from multiple views. We have evaluated our method on varying plant structures and compared it with more traditional methods using arbitrary camera positions and static cameras, in terms of the number of points obtained and the accuracy of these with respect to the Euclidean distance to the ground truth.

In all experiments, our AVC produces more data of higher accuracy, with a reduced image set. More points help ensure that the plant has been scanned adequately and that the amount of unknown object data are minimal. More accurate points ensure that the 3D model can be reconstructed with increased fidelity, which is vital for accurate plant phenotyping. The AVC acquires more points per image, indicating that the images captured provide more value toward reconstruction. While static camera placement can be effective, there are clear data gains to be made by employing active vision.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. 3D reconstructions generated by the comparable imaging methods.

Received June 4, 2018; accepted July 27, 2018; published August 10, 2018.

LITERATURE CITED

Adeloye A (2010) Global warming impact: flood events, wet-dry conditions and changing scene in world food security. *J Agric Res Dev* 9:

Aloimonos J, Weiss I, Bandyopadhyay A (1988) Active vision. *Int J Comput Vis* 1: 333–356

Burgess AJ, Retkute R, Pound MP, Foulkes J, Preston SP, Jensen OE, Pridmore TP, Murchie EH (2015) High-resolution three-dimensional structural data

quantify the impact of photoinhibition on long-term carbon gain in wheat canopies in the field. *Plant Physiol* **169**: 1192–1204

Burgess AJ, Retkute R, Preston SP, Jensen OE, Pound MP, Pridmore TP, Murchie EH (2016) The 4-dimensional plant: effects of wind-induced canopy movement on light fluctuations and photosynthesis. *Front Plant Sci* **7**: 1392

Burgess AJ, Retkute R, Herman T, Murchie EH (2017) Exploring relationships between canopy architecture, light distribution, and photosynthesis in contrasting rice genotypes using 3D canopy reconstruction. *Front Plant Sci* **8**: 734

Challinor A, Watson J, Lobell DB, Howden SM, Smith DR, Chhetri N (2014) A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change* **4**: 287–291

Dornaika F, Horraud R (1998) Simultaneous robot-world and hand-eye calibration. *IEEE Trans Robot Autom* **14**: 617–622

Faaij A (2008) Bioenergy and Global Food Security. http://www.wbgu.de/fileadmin/templates/dateien/veroeffentlichungen/hauptgutachten/jg2008/wbgu_jg2008_ex03.pdf

Furbank RT, Tester M (2011) Phenomics: technologies to relieve the phenotyping bottleneck. *Trends Plant Sci* **16**: 635–644

Furukawa Y, Ponce J (2010) Accurate, dense, and robust multiview stereopsis. *IEEE Trans Pattern Anal Mach Intell* **32**: 1362–1376

Gibbs JA, Pound M, French AP, Wells DM, Murchie E, Pridmore T (2017) Approaches to three-dimensional reconstruction of plant shoot topology and geometry. *Funct Plant Biol* **44**: 62

Hemming J, Bac CW, van Tuyl BAJ, Barth R, Bontsema J, Pekkeriet E, van Henten EJ (2014a) A robot for harvesting sweet-pepper in greenhouses. In: *Proceedings of the International Conference of Agricultural Engineering*. <http://edepot.wur.nl/309949>

Hemming J, Ruizendaal J, Hofstee JW, van Henten EJ (2014b) Fruit detectability analysis for different camera positions in sweet-pepper. *Sensors (Basel)* **14**: 6032–6044

Hornung A, Zeng B, Kobbelt L (2008) Image selection for improved multi-view stereo. In: *Computer Vision and Pattern Recognition, 2008: IEEE Conference on CVPR 2008*. IEEE, pp 1–8

Long SP, Zhu XG, Naidu SL, Ort DR (2006) Can improvement in photosynthesis increase crop yields? *Plant Cell Environ* **29**: 315–330

Massios NA, Fisher RB (1998) A best next view selection algorithm incorporating a quality criterion. In: *Proceedings of the British Machine Vision Conference BMVC98*. pp 780–789

Paproki A, Sirault X, Berry S, Furbank R, Fripp J (2012) A novel mesh processing based technique for 3D plant analysis. *BMC Plant Biol* **12**: 63

Poorter H, Fiorani F, Pieruschka R, Wojciechowski T, van der Putten WH, Kleyer M, Schurr U, Postma J (2016) Pampered inside, pestered outside? Differences and similarities between plants growing in controlled conditions and in the field. *New Phytol* **212**: 838–855

Pound MP, French AP, Murchie EH, Pridmore TP (2014) Automated recovery of three-dimensional models of plant shoots from multiple color images. *Plant Physiol* **166**: 1688–1698

Quan L, Tan P, Zeng G, Yuan L, Wang J, Kang SB (2006) Image-based plant modeling. *ACM Trans Graph* **25**: 599

Remondino F, El-Hakim S (2006) Image-based 3D modelling: a review. *Photogramm Rec* **21**: 269–291

Reynolds MP, van Ginkel M, Ribaut JM (2000) Avenues for genetic modification of radiation use efficiency in wheat. *J Exp Bot* **51**: 459–473

Seitz SM (2000) A theory of shape by space carving. *Int J Comput Vis* **38**: 199–218

Seitz SM, Curless B, Diebel J, Scharstein D, Szeliski R (2006) A comparison and evaluation of multi-view stereo reconstruction algorithms. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol 1. IEEE, pp 519–528

Sticklen MB (2007) Feedstock crop genetic engineering for alcohol fuels. *Crop Sci* **47**: 2238

Vasquez-Gomez JI, Sucar LE, Murrieta-Cid R (2013) Hierarchical ray tracing for fast volumetric next-best-view planning. In: *2013 International Conference on Computer and Robot Vision*. IEEE, pp 181–187

Wong LM, Dumont C, Abidi MA (1999) Next best view system in a 3D object modeling task. In: *Proceedings 1999 IEEE International Symposium on Computational Intelligence in Robotics and Automation*. IEEE, pp 306–311