Three-Dimensional Reconstruction of Plant Shoots from Multiple Images using an Active Vision System

Jonathon A. Gibbs, Michael Pound, Darren M. Wells, Erik Murchie, Andrew French, Tony Pridmore

Abstract **–** The reconstruction of 3D models of plant shoots is a challenging problem central to the emerging discipline of plant phenomics – the quantitative measurement of plant structure and function. Current approaches are, however, often limited by the use of static cameras. We propose an automated active phenotyping cell to reconstruct plant shoots from multiple images using a turntable capable of rotating 360 degrees and camera mounted robot arm. To overcome the problem of static camera positions we develop an algorithm capable of analysing the environment and determining viewpoints from which to capture initial images suitable for use by a structure from motion technique.

I. Background and Motivation

The global population is expected to reach 9 billion by 2050 and the spread of prosperity throughout the world is increasing the food intake per capita, driving the demand for a richer, more varied diet. At the same time, changes in climate are causing more frequent and severe flooding, destroying crop yields and shortage of arable land constitutes an additional challenge. It has been widely predicted that without crop climate adaption the production of food will deteriorate [1], [2]. The long-term goal of this work is to provide the innovative approach to sustainable agriculture necessary to adapt to the fluctuating environment and increased demand for food.

The identification of more productive and/or resilient crop species requires connections to be made between the genetic and physical structures of the plant. While significant progress has been made in the study of the genome in recent years, the creation and quantitative analysis of plant phenotypes (structures) has become a major bottleneck. Though some plant traits (e.g. leaf area) can be estimated using a single carefully placed camera and 2D image analysis methods, the ability to produce accurate 3D models of plants would support a wide variety of phenotyping tasks.

Image-based reconstruction methods are attractive in this context. Plants are easily disturbed; non-invasive sensing techniques capable of capturing information across the whole object are required. Plant shoots are, however, a challenging target for image-based reconstruction. Individual variation within species is often large, making it difficult to predict structures a priori. Individual leaves can be very similar in appearance, and densely-packed, occluding each other from many viewpoints: plants can be very crowded scenes. The leaves of many species are quite highly reflective, and often lack the strong texture needed by some techniques.

The starting point for the work described here is the hypothesis that active vision can aid in the generation of high-quality plant models by providing improved, and responsive, image acquisition strategies. Active vision systems automatically control and manipulate camera viewpoints to provide images which best support the task at hand. Active methods have played a role in other plant-related tasks. For example, [3] attach a camera to a robot arm in order to identify peppers to be collected. The effect of camera placement on fruit picking has been investigated [4], and active vision used to address the problem of occlusion. The large-scale phenotyping systems now finding application in plant and crop science, however, typically rely on fixed viewpoints that are not adapted to the specific plant being modelled. Some systems rotate the plant during imaging, but still use static camera positions. This means that, in many cases, the images captured are far from optimal, adversely affecting the results obtained. The ability to adjust sensors in response to emerging plant properties (e.g. size) is vital if accurate representations are to be obtained of a wide variety of plant species, ages and conditions.

We aim to produce a fully automated, active system 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 interaction, with the longer-term goal of improved plant phenotyping. The approach proposed here offers more flexibility than existing large scale phenotyping systems by adapting to the natural variation of individual plants in order to obtain optimal data.

The remainder of the paper is organized as follows; we first introduce the reader to 3D plant reconstruction, discussing current approaches and the challenges they face. We then provide a concise overview of active vision and the various components that are necessary, before discussing the approach used in this work. Results obtained from real and artificial plants are presented. Finally, we conclude with a summary of progress and plans for future work.

II. 3D Plant Reconstruction

Until the late 1960s botanical drawings were the primary means of capturing plant architecture. Today a variety of approaches are available. Rule-based methods use a set of rules to define the structure of a particular species or class of plant. Varying the parameters of these systems produces models of single plants, but rule-based approaches cannot easily be used to produce the descriptions of specific, existing plants needed to support phenotyping.

Image based approaches seek plant geometry directly, analysing a set of images to reconstruct representations of actual plants. Image based models can be used to support simulations and enable the extraction of trait measurements.

Some approaches, such as Light Detection and Ranging (LiDAR) [5], custom illuminate the target object by emitting radiation into the scene. LIDAR is commonly used in the airborne reconstruction of field based plants, trees in particular. For example, [6] describe the forest canopy as a series of cones fitted to a raw LiDAR point cloud, then apply simple geometric operations to adjust and correct its height. Similar methods can be applied to smaller plants; [7] model rice plants using a three-dimensional sonic digitiser to capture a 3D point cloud. The digitisation process is reported to take up to an hour to complete, and [8] note that the digitisation process for their approach to reconstruct White Clover canopies required between 3 to 7 hours. They used electromagnetic digitising apparatus with corner flags to aid calibration, applying a destructive approach and pruning the canopy from the top downwards.

The recovery of 3D descriptions from images captured under natural illumination is a longstanding research topic in the computer vision community. A range of approaches such as structure from motion, shape-from-silhouette and space carving, have been developed and can be used for plant reconstruction. For example [9] combine a volumetric opacity estimate with view-dependent texturing and successfully model trees from a series of images whilst [10] use a space carving approach with particle flows to estimate tree volume. [13] use a stereovision approach to reconstruct plant models using automated segmentation. User input is, however, often required. [11] adopt the less common approach of sketching to create plant models. Other interactive approaches construct models directly from images. [12] obtain a point cloud from 35 images of a plant, though user input is required in the form of segmentation to separate leaves, and the image acquisition process is manual.

Fully automatic reconstruction of plants from natural images is challenging due to the intricate phyllotaxis (leaf structure) and continuous reorganization of plant foliage. Many problems arise during the image acquisition and reconstruction processes. Determining the number of images required, and their viewpoints, such that all the required plant features are visible remains difficult. Too few or poorly chosen images results in the loss of data, whilst too many results in increased computational requirements.

Occlusions are a common side effect of complex structures such as plants and can be overcome by capturing an increased number of images, though in some cases approximation techniques must be used. Some approaches rely on intrusive/destructive approaches to obtain more information, however this means the plant cannot return to its original configuration, preventing the comparison of descriptions obtained at different times. Invasive methods can also increase reconstruction time and encourage irreversible errors. Multiple side image methods also exist but often don't support 3D modelling as there is no overlap between images.

An active vision approach can alleviate the problems associated with plant modelling. By manipulating the camera(s) to optimise image number and viewpoint it can help overcome occlusion. By analysing a developing point cloud and moving to view a region that has been identified as unexpectedly sparse, it can help to obtain missing data. Selecting camera positions on the basis of emerging data can also prevent multiple, unnecessary views of the same regions being collected, both reducing the computational requirements and explicitly reacting to natural variation.

III. An Active Phenotyping Cell

A. *Hardware and Calibration*

We present a nonintrusive and nondestructive active vision approach to 3D plant modelling using a camera mounted robot arm and a turntable. The approach is based on a structure from motion method that derives 3D descriptions of the plant surface from sets of colour images. Our active phenotyping cell comprises a Universal Robot 5 (UR5), with a standard handheld camera, Canon 650D, and a high precision turntable, the LT360 EX. The UR5 offers 6 degrees of freedom whilst the turntable enables a full 360 degrees of rotation ensuring it is possible to see the entire plant, both of which are necessary as it is not always possible for the robot arm to move around the entire plant, for example a large rice plant. Our setup is illustrated in Figure 1.

Figure 1. Hardware setup of Robot, Turntable and Camera

We calibrate the camera using a checkerboard approach [14], in which 15 arbitrary images of the checkerboard are captured. We calculate the forward kinematics using Denavit Hartenberg (DH) parameters [15] with joint angles obtained directly from the robot. The remaining transformations are calculated using a simultaneous closed-form quaternion approach [16].

In order to use the turntable with our fully calibrated system we need to take into consideration the rotations performed by the turntable. To achieve this we project to the centre of the turntable, which is known from our calibration process. From this we can calculate Y_j where Yr, the rotation of Y, is calculated using Eq. 1, j is the number of degrees that the turntable has rotated and function RotZ is a rotation around the Z axis.

$$\mathrm{YR_j} = \left(\mathrm{Y}' \mathrm{RotZ(j)}\right)' \qquad \text{(Eq. 1)}$$

The translation, Yt, requires that the difference between the rotation matrix before and after a rotation is known; Eq.

2, where $\vec{c}$ is the homogeneous position of the centre of the turntable. Finally, we multiply x' and y' by Yr with its original translation from Y_0 to obtain Y_j

$$\vec{p} = \vec{c} - (RotZ(j)\vec{c})$$

$$x' = -\vec{c_x} + (-\vec{p_x})$$

$$y' = -\vec{c_y} + (\vec{p_y})$$

$$\vec{Yt_j} = [x' y' 0\ 1]' \qquad \text{(Eq. 2)}$$

Y_j can then be calculated as

$$Y_j = [YR_j \vec{Yt_0}] * \vec{Yt_j} \qquad \text{(Eq. 3)}$$

Once we have a fully calibrated system we are able to remove the checkerboard from the scene and calculate our camera position from the remaining variables.

B. *Image acquisition strategies*

To obtain accurate 3D models via structure from motion the camera needs to be in a position to collect an optimal number of images of the highest quality. This is a challenging problem due to the vast number of possible viewpoints and the lack of prior knowledge of the shape and size of the object. We have developed a proof-of-concept image acquisition strategy that uses a simple threshold-based method to identify the plant in order to calculate initial camera positions. There are two primary constraints; 1. The camera must be facing the plant in the robot's starting position, approximately placing the plant in the centre of the view. 2. A white background must be used with no other colour visible, which allows us to calculate the position of any given plant. These constraints are commonly satisfied and/or are easily achievable in controlled phenotyping environments. More powerful segmentation methods could be used in less constrained environments.

The role of image analysis in the proposed system is to identify four points on the boundary of the plant region; those nearest the four edges of the image. The coordinates of these points provide measures, TX, BX, LX, RX, of the shortest distances from the plant region to top, bottom left and right edges respectively. A user-defined threshold is applied to separate plant from (white) background, and plant pixels with the highest and lowest x and y coordinates are identified. To reduce the likelihood of selecting a noise-generated false-positive plant pixel we examine 400 pixels around each candidate (approximately 0.01% of the total pixels). Only if 75% or more are of those pixels are above threshold is the pixel accepted as lying on and near the boundary of the plant (Figure 2). This heuristic is simple, but effective and computationally efficient.

To initialise and parameterise the system the camera is first moved to a start position facing the turntable. It is then moved in a plane normal to the image plane to define four points. These points define the corners of a quadrilateral normal to the image and passing through the start position. The points are chosen to be the furthest from the start point in each direction from which the turntable remains visible. Throughout image acquisition all translational movements of the camera take place within the plane defined by this quadrilateral. Camera rotations may take it outside the plane, but it remains close to it at all times.

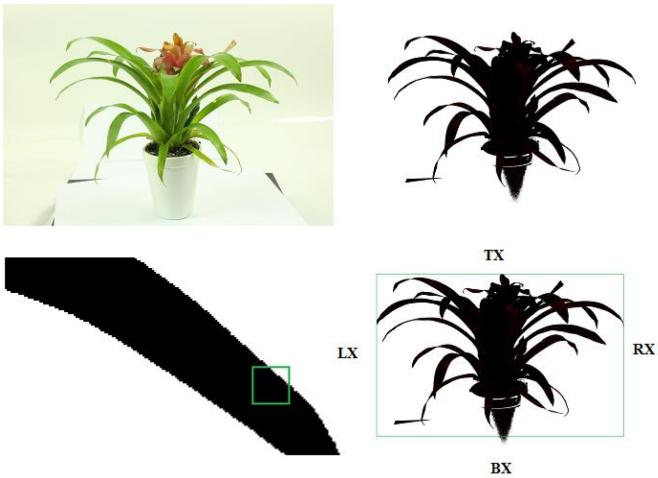

FIGURE 2. TOP LEFT: ORIGINAL IMAGE, TOP RIGHT: IMAGE SEPARATED FROM BACKGROUND WITH DISCARDED OUTLIERS, BOTTOM LEFT: EVALUATION OF A PIXEL, BOTTOM RIGHT: THE RESULTING BORDER DEFINING VALUES RX, LX, TX, BX

The centre of the quadrilateral is used to define a set of *n* initial points from which the search for suitable viewpoints begins. These are evenly spaced along a vertical line through the quadrilateral centre; the image acquisition process is run from these points in fixed (lowest to highest) order, providing *n* images for each turntable position. The dimensions of the quadrilateral determine the size of the camera translations made during image acquisition. Large translations towards (forward) or away from the plant (backwards) are 30% of the width of the quadrilateral, small movements 10%. Camera rotations (up, down, left, right) are of a small, fixed size (typically 2 deg.) set by the user.

Active image acquisition begins with the camera in one of the initial positions described above. Images are repeatedly captured, thresholded, plant boundary points identified, and the camera moved under the control of a set of heuristic rules until the plant is either fully enclosed by the image boundary but without excess space or the arm is at its maximum reach.

The rules employed are intuitive, but effective:

- If there are 50 or more pixels of white space surrounding the plant (TX, LX, BX, RX all > 50) a forward movement is made.
- If the plant region is close to the boundary at either the top and bottom or left and right a backward movement is made.
- Forward and backwards movements are large unless a movement in the opposite direction has just been made, in which case they are small. This introduces a degree of fine-tuning and prevents oscillation.

- If LX is large and RX is small, rotate left.
- If RX is large and LX is large, rotate right.
- If TX is large and BX is small, rotate upwards.
- If TB is large and TX is small, rotate downwards.

These rules are applied to each of the vertical stack of initial points. Once an improved camera position has been identified for each such point, images are captured and the turntable is rotated. The size of the rotation is set to ensure that at least 60 images are captured in total. In a typical experiment 6 vertical positions are used, and the turntable rotated 36 degrees, between capture sessions. During image capture camera files are created containing the camera matrix that transforms a 3D point to a 2D point on the image plane.

Plant structure varies significantly between species; when modelling those expected to be rotationally symmetric the search for camera locations need only be performed once and the same positions used at each turntable rotation. Given species that may not be rotationally symmetric a new search may be performed for each turntable setting.

C. Reconstruction methods

A point cloud is first generated from the images and corresponding transformation matrices using Patch-based Multi-view Stereo [17]. The point cloud is the starting point for further reconstruction and is a common input for many software packages and surface reconstruction algorithms.

We also apply Pound et al's Canopy Reconstruction method [18] which accepts a point cloud as input and generates a surface using alpha-shapes and level set methods, aiding the process by revisiting the images to ensure consistency. Note that this final stage is not possible when using a direct 3D sensor such as a laser scanner.

Surface reconstruction is fully automated and only requires user interaction if the hardware is moved, in which case the calibration stage needs to be performed. Patch-Based Multi-View Stereo (PMVS) and Canopy Reconstruction have been integrated into our cell to create a smooth workflow that can run unattended, taking a step towards reducing the phenotyping bottleneck. PMVS takes a set of images and camera parameters and reconstructs the 3D structure and Canopy Reconstruction takes the output to generate a surface-based description.

IV. RESULTS AND DISCUSSION

We conducted experiments on four artificial plants of varying sizes and densities (Figure 3). Models of each plant were built using a set of fixed camera positions, defined such that the largest of our plants is fully visible in each image, and results compared to those obtained from our active vision system, which reacts to the size of the plant. A set of 60 images were used for each reconstruction.

This initial study focuses first on the point cloud data provided by PMVS. Comparison of the number of high quality points generated from static and actively captured images by this state of the art method gives some insight into the potential benefits of the active approach. Figure 4 shows the point clouds obtained from each image set, for clarity we have manually removed, using Meshlab, the excess data obtained from under the plant, mainly from the plant pot.

FIGURE 3. ORIGINAL PLANTS, TOP ROW PLANTS A AND B, SECOND ROW PLANTS C AND D

We compare the number of points obtained by static and active vision for each plant; Plant A active 120,422 whereas static produces significantly less at 35,872. Plant C active has 99,570 points compared to 26,668 static and Plant D active 51,267 points and 17,388 static. The static camera positions were in fact obtained by running the active method over plant B, ensuring that the largest plant is fully visible in the images and therefore has the same number of points for both static and active; 168,344. Active vision provides significantly more valuable points for each plant, which is particularly useful for the small dense plants in this study.

Though point clouds capture the broad structure of the target object, surface reconstruction is essential for plant phenotyping, as many desirable traits must be measured over leaves. The canopy reconstruction method of [18] was applied to the actively acquired image sets generated here; results are shown in Figure 5. Our artificial test plants are particularly challenging, with only very small 3D and colour differences between their very close packed, uniform leaves. [18] employs an image-based surface patch extension method which produced an acceptable surface reconstruction, but tended to over-extend leaves. We applied the same techniques to a dense domestic plant (Figure 6). The noise present in the point cloud (middle) has successfully been removed by [18] which uses a colour threshold to remove noisy points. The point clouds and images obtained from our initial active phenotyping cell can support fully automatic 3D modelling of real plants.

More complex reconstruction algorithms such as [18] may also benefit from the integration of active image acquisition strategies, but have different requirements than point cloud recovery methods. Though [18] builds on data supplied by PMVS, choosing images to simply increase the number of points may not be the best strategy. [18] operates within planar patches fitted to point clusters – increasing the number of points available need not improve the plane descriptions, and could add noise. The points provided by PMVS arise from textured leaf areas; [18] may benefit more from strategies that provide clearer views and higher resolution images of smoother (less textured) areas, allowing a greater degree of patch extension while making leaf boundaries more easily identifiable. This could be achieved by exploiting initial surface reconstruction data to guide acquisition of new images, rather than selecting them from a pre-acquired set as is currently the approach in [18].

FIGURE 4. LEFT HAND SIDE ACTIVE VISION POINT CLOUDS, RIGHT SIDE STATIC. THE PLANTS FROM TOP TO BOTTOM ARE A, B, C, D

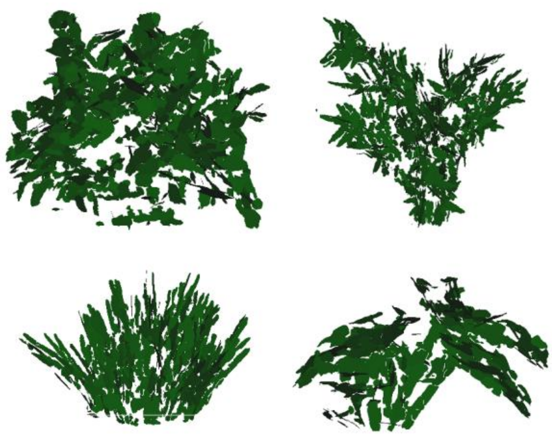

FIGURE 5. SURFACE RECONSTRUCTION FOR THE FOUR ARTIFICIAL PLANTS CORRESPONDING TO THOSE SHOWN IN FIGURE 3

Point cloud data may also be used to guide image acquisition. Though point clouds provide a relatively crude representation of complex plant architectures, they can capture plants' broad structure. Models of the expected distribution of points in different species might highlight regions of the target volume that are not sufficiently explored by an initial image set, allowing the camera to viewpoints that will produce more complete plant descriptions.

FIGURE 6. SURFACE RECONSTRUCTION OF A REAL PLANT FROM ACTIVELY ACQUIRED IMAGES. TOP; ACTUAL PLANT, MIDDLE; THE POINT CLOUD ACQUIRED, BOTTOM; SURFACE RECONSTRUCTION

V. CONCLUSION

We present initial work towards an active plant phenotyping cell capable of recovering 3D descriptions of plant shoots from multiple colour images. An automatic image acquisition technique is described which provides improved point cloud data and supports the 3D reconstruction of leaf surfaces. After the initial calibration of the system, which need only be done if hardware is replaced or moved, no user input is necessary and the process can continuously run through a custom designed interface. Experimental results show that by using an active vision approach, rather than a traditional static set of camera positions we are able to gather significantly more data on the plant and its structure from the same number of images.

The active vision approach provides significant opportunities to enhance and extend the scope of surface reconstruction methods such as [18]. Careful selection of views focusing on areas of ambiguity will, we believe, produce both more accurate point clouds and higher quality image data from which surfaces can be produced. Active vision may also reduce the number of unnecessary images captured, those adding little to the reconstruction, improving throughput.

Cameras and multi-view stereo are employed here, rather than e.g. laser scanners, as the image sets involved carry information on plant appearance missing from a point cloud. In addition to providing 3D structure, multiple colour images could be used e.g. to assess plant health. We would suggest, however, that an active sensing approach could aid the integration of the 2½D data produced by such devices.

Future work will more closely integrate active image acquisition into the reconstruction process, allowing a wide range of camera movements and focusing on areas of ambiguity, occlusion and those likely to be missing data. Evaluation of the models produced is difficult as ground truth data is required. Future work will also investigate the possibility of using X-ray CT data to produce reference data. In the longer term we aim to provide improved, active phenotyping of a wide range of complex plant species.

REFERENCES

[1] A. Challinor, J. Watson, D. Lobell, S. Howden, D. Smith, and N. Chhetri, "A meta-analysis of crop yield under climate change and adaptation," *Nature Climate Change*. Nature Publishing Group, 16-Mar-2014.

[2] A. Adeloye, "Global Warming Impact: Flood Events, Wet-Dry Conditions and Changing Scene in World Food Security," *J. Ag.. Res. Dev.*, 9, 1, 2010.

[3] J. Hemming, C. W. Bac, B. A. J. van Tuijl, R. Barth, J. Bontsema, E. J. Pekkeriet, and E. van Henten, "A robot for harvesting sweet-pepper in greenhouses," in *Proceedings of the International Conference of Agricultural Engineering*, 2014.

[4] J. Hemming, J. Ruizendaal, J. W. Hofstee, and E. J. van Henten, "Fruit detectability analysis for different camera positions in sweet-pepper.," *Sensors (Basel).*, vol. 14, no. 4, pp. 6032–44, Jan. 2014.

[5] C. A. Northend, "Lidar, a laser radar for meteorological studies," *Naturwissenschaften*, vol. 54, no. 4, pp. 77–80, 1967.

[6] M. Van Leeuwen, N. C. Coops, and M. A. Wulder, "Canopy surface reconstruction from a LiDAR point cloud using Hough transform," *Remote Sens. Lett.*, vol. 1, no. 3, pp. 125–132, Sep. 2010.

[7] T. Watanabe, J. S. Hanan, P. M. Room, T. Hasegawa, H. Nakagawa, and W. Takahashi, "Rice morphogenesis and plant architecture: measurement, specification and the reconstruction of structural development by 3D architectural modelling.," *Ann. Bot.*, vol. 95, no. 7, pp. 1131–43, Jun. 2005.

[8] M. Rakocevic, "Assessing the Geometric Structure of a White Clover (Trifolium repens L.) Canopy using3-D Digitising," *Ann. Bot.*, vol. 86, no. 3, pp. 519–526, Sep. 2000.

[9] A. Reche-Martinez, I. Martin, and G. Drettakis, "Volumetric Reconstruction and Interactive Rendering of Trees from Photographs," in *ACM Transactions on Graphics (ToG)*, 2004, vol. 23, no. 3, pp. 720 – 727.

[10] B. Neubert, T. Franken, and O. Deussen, "Approximate image-based tree-modeling using particle flows," in *ACM SIGGRAPH 2007 papers on - SIGGRAPH '07*, 2007, vol. 26, no. 3, p. 88.

[11] M. Masry and H. Lipson, "A sketch-based interface for iterative design and analysis of 3D objects," in *ACM SIGGRAPH 2007 courses on - SIGGRAPH '07*, 2007, p. 31.

[12] P. Tan, L. Yuan, and J. Wang, "Image-based Plant Modeling Overview of Plant Modeling System," pp. 599–604, 2003.

[13] B. Biskup, H. Scharr, U. Schurr, and U. Rascher, "A stereo imaging system for measuring structural parameters of plant canopies.," *Plant. Cell Environ.*, vol. 30, no. 10, pp. 1299–308, Oct. 2007.

[14] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330 – 1334, 2000.

[15] J. Denavit, "A kinematic notation for lower-pair mechanisms based on matrices.," *J. Appl. Mech.*, vol. 20, pp. 215–222, 1955.

[16] F. Dornaika and R. Horaud, "Simultaneous robot-world and hand-eye calibration," *IEEE Trans. Robot. Autom.*, vol. 14, no. 4, pp. 617–622, 1998.

[17] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis.," *IEEE Trans. PAMI.*, vol. 32, no. 8, pp. 1362–76, Aug. 2010.

[18] M. P. Pound, A. P. French, E. H. Murchie, and T. P. Pridmore, "Automated recovery of three-dimensional models of plant shoots from multiple color images.," *Plant Physiol.*, vol. 166, no. 4, pp. 1688–98, Dec. 2014.