Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/29219/1/Binder1.pdf

Running head: Automated recovery of 3D models of plant shoots

Corresponding Author: Tony Pridmore

Centre for Plant Integrative Biology, School of Biosciences, University of Nottingham, Sutton Bonington, LE12 5RD, UK.

Tel: +44 (0)115 8466510

Email: tony.pridmore@nottingham.ac.uk

Research Area: Breakthrough Technologies

Title: Automated recovery of 3D models of plant shoots from multiple colour images

Authors: Michael P Pound[1], Andrew P French[1,2], Erik H Murchie[1], Tony P Pridmore[1,2]

[1]Centre for Plant Integrative Biology, School of Biosciences, University of Nottingham, Sutton Bonington, LE12 5RD, UK.

[2]School of Computer Science, University of Nottingham, Jubilee Campus, Wollaton Rd, Nottingham, NG8 1BB, UK

Summary: A fully automatic approach to 3D plant shoot reconstruction using multiple images taken with a single camera.

Corresponding Author: Tony Pridmore (tony.Pridmore@nottingham.ac.uk)

Abstract

Increased adoption of the systems approach to biological research has focussed attention on the use of quantitative models of biological objects. This includes a need for realistic 3D representations of plant shoots for quantification and modelling. Previous limitations in single or multi-view stereo algorithms have led to a reliance on volumetric methods or expensive hardware to record plant structure. We present a fully automatic approach to image-based 3D plant reconstruction that can be achieved using a single low-cost camera. The reconstructed plants are represented as a series of small planar sections that together model the more complex architecture of the leaf surfaces. The boundary of each leaf patch is refined using the level set method, optimising the model based on image information, curvature constraints and the position of neighbouring surfaces. The reconstruction process makes few assumptions about the nature of the plant material being reconstructed, and as such is applicable to a wide variety of plant species and topologies, and can be extended to canopy-scale imaging. We demonstrate the effectiveness of our approach on datasets of wheat and rice plants, as well as a novel virtual dataset that allows us to compute quantitative measures of reconstruction accuracy. The output is a 3D mesh structure that is suitable for modelling applications, in a format that can be imported in the majority of 3D graphics and software packages.

Introduction

In recent years there has been a surge in interest in the construction of geometrically accurate models of plants. Increased adoption of the systems approach to biological research has focussed attention on the use of quantitative models of biological objects and processes to both make and test hypotheses. Applications of the systems approach are diverse, and include the study of canopy photosynthesis (e.g. Song et al. 2013, Watanabe 2005). The work reported here was motivated by a need for realistic 3D representations of plant shoots for photosynthesis modelling. Canopy structure or 'architecture' is an important co-determinant of the maximum productivity of crops and theoretically can influence canopy photosynthesis efficiency (Zhu et al. 2004, Reynolds et al. 2013, Long et al. 2006). For example many high yielding cereal varieties tend to have upright leaves which raises the optimal leaf area index and reduces the proportion of leaves in a light-saturated condition (Murchie and Reynolds 2012). However there is still considerable variation in plant canopy architecture and it is currently unclear whether many existing architectures (which have been influenced by breeding for specific traits) offer optimal arrangements for photosynthesis (Long et al. 2006, Murchie et al. 2009). Therefore there is a need for rapid, automated, user-accessible techniques for accurately measuring 3D architecture of complex crop canopies.

Existing plant modelling approaches can be broadly classified as either rule-based, or image-based (Quan et al. 2006). Rule-based approaches generate model plants based on rules or grammars with specified structure. These rules, and hence the form and parameters of the models produced, are often derived from measurements of real plants (e.g. Watanabe, 2005. Alarcon et al., 2011). The resulting virtual plants can model different phenotypes, plant response to various growing conditions and stresses, and when based on real-world data will be reasonably accurate. However, the data acquisition process is often extremely time consuming, and is usually tailored to a particular species. In many cases only a small set of varieties can be described, due to the manual measurements required to parameterise the model. It is also difficult to incorporate additional data into a rule-based model after it has been created. For example, modelling the effect of environmental factors on plant development will require rules to be modified and additional measures be taken, again a time-consuming process. Critically, while models created in this way may capture the important characteristics of a given species, they do not necessarily describe any specific, individual plant. This is important when considering the variation in plant structure that can occur in response to environment and features that are associated with communities of plants, including crop canopies. This can be seen in the relationship between planting density and tillering rate, for example (Zhong et al, 2002).

Image-based approaches attempt to directly model a given object by extracting the necessary information from one or more images of that object. The approach relies on techniques developed in the wider field of computer vision, and is computationally challenging. The continued increase in available computing power, however, means this can be done efficiently, and in some cases fully automatically. Image-based approaches to plant modelling are particularly attractive as, in addition to supporting systems biology, they provide a route to plant phenotyping (Houle et al. 2010, White et al. 2012). Three-dimensional geometrical models contain the information needed to compute summary plant traits, such as total leaf area, mean leaf angle, etc. These underpin both plant breeding programmes and attempts to understand the relationship between genotype, phenotype and environment, regardless of the scientific approach taken.

Plants, however, provide a particularly challenging image-based modelling target, including large amounts of self-occlusion (leaves obscuring one another), and many smaller surfaces

that appear similar. Depending on plant species, leaves can lack the texture necessary to perform robust feature matching, either to separate leaves from one another, or locate specific leaves across multiple views. To overcome this, where image-based modelling approaches are successful, they have often involved user-interaction to guide the process (Quan, 2006).

It is possible to further categorise image-based approaches into those that are top-down, beginning with an existing, generic, plant model that is fitted to the received image data, and those that are bottom-up, creating a description of the plant by examining the contents of the images. Top-down approaches attempt to simplify the model construction problem by instead solving a model refinement problem. An existing model is adjusted to fit the image data, so that the new plant representation is consistent with what is observed. Quan (2006) takes this approach, first obtaining an ideal leaf model from a single leaf, and then fitting it to all other (user-segmented) leaves in the scene. By adapting an existing model, topological inconsistency (such as the self-intersection of leaf surfaces) is avoided, but this comes at the expense of generality. This approach can be extended in principle to other plant species, but where the geometry of a plant or leaf differs greatly from the expected model, the suitability of this approach is unclear. This is also significant where a plant model has been generated with highly specific real-world data, such as the work of Watanabe (2005).

Bottom-up methods begin with one or more images, and reconstruct a plant model based only on the observed pixel data. Two broad approaches exist, both requiring a set of images captured from different, but known, viewpoints. Silhouette-based methods (Clark, 2011. Kumar, 2012) segment each image to identify the boundary of the object of interest. Each silhouette is effectively projected into the viewed environment, identifying a region of 3D space that could be occupied by the plant. These regions are combined to determine the maximum possible object size that is consistent with the images presented to the algorithm. In many cases, where the number of input images is high, the resulting model will be a good approximation to the true plant structure. However, as the scene becomes increasingly complex, for example with the addition of more leaves in an older plant, the discrepancy between true object and model will increase. This problem becomes more pronounced when extending these techniques to very complex scenes such as plant canopies, where its effectiveness is limited.

Correspondence-based methods identify features of interest (i.e. recognisable patches of pixels in the image), independently in each of a set of images, and then match those features between views. If the image features associated with a particular plant feature (e.g. the tip of a leaf) can be identified in multiple images taken from different viewpoints, knowledge of the cameras' positions and orientations allow its 3D location to be computed. The result is a point cloud representation, a set of (x,y,z) coordinates approximating the plant's surface (Quan, 2006. Omasa, 2007. Alenya, 2011). Point cloud data can be obtained directly from special purpose hardware devices such as LIDAR (Omasa 2007) and time-of-flight laser (Alenya, 2011) scanners. The equipment involved, however, can be expensive, and often places restrictions on the environments in which it can operate.

Image-based modelling algorithms are widely applicable and require only easily accessible and affordable cameras. Their generality can, however, become a hindrance, as the challenging nature of plant topology may require additional assumptions to be made as the reconstruction proceeds. The representations they produce may also be unsuitable for direct use in some situations. The volumetric data structures produced by silhouette-based methods, for example, are static: the size and position of the voxels are defined early in the process and are difficult to change. While measurements of e.g. height and volume are easily made from volumetric descriptions, estimating motion e.g. of leaves moving in the breeze is extremely

difficult. Similarly, point clouds can be used to calculate density and distributions of plant material, but cannot immediately be used to model photosynthetic activity: for this a surface-based representation is required.

This paper describes a fully automatic, bottom-up approach to image-based 3D plant reconstruction that is applicable to a wide variety of plant species and topologies. The method is accurate, providing a true representation of the original plant, and produces data in a form that can support both trait measurement and modelling techniques such as forward ray tracing (Song et al., 2013).

The proposed approach requires a point cloud and a set of colour images. We obtain these using existing correspondence-based techniques (Furukawa, 2010. Wu, 2011), with a focus on inexpensive equipment, however hardware-based approaches that generate point clouds in 3D, such as LIDAR scanners, could also be used. The point cloud is first described by a set of planar patches, each representing a small section of plant material, usually a segment of leaf. Where the quality of the input point cloud is high, the initial surface estimate will provide a good model of the plant. Image noise and the complexity of the plant will, however, typically lead to missing areas of leaf material, and poorly defined leaf boundaries. We therefore extend existing approaches by refining the initial surface estimate into a more accurate plant model. Initial surface patches are re-sized and re-shaped based on image information, and information obtained from neighbouring surfaces. The resulting surface patches are then re-triangulated to produce a smooth and geometrically accurate model of the plant.

The reconstruction process makes few assumptions about the nature of the plant material being reconstructed; by representing each leaf as a series of small planar sections, the complete leaf surface itself can take any reasonable shape. The output is a 3D mesh structure that is suitable for ray-tracing or other modelling applications, in a format that can be imported in the majority of 3D graphics and software packages. The generality of our technique allows it to be scaled to scenes involving multiple plants, and even plant canopies. However, the focus of this paper is on the accurate reconstruction of single plants of varying species. In the discussion section we will outline our intentions for extending this method to field-scale phenotyping.

A software tool utilizing the techniques described in this paper has been released as an open-source project under BSD license.

Results

Growth of plants

Cultivation of wheat plants took place in glasshouses on Sutton Bonington campus, University of Nottingham during the summer of 2013. Seeds of the variety 'Paragon' were germinated on Levington's seed and modular compost. Following vernalisation they were transferred to 2L pots with J. Arthur Bower's John Innes #2 soil-based compost. Plants were watered daily and natural light as supplemented with sonTAgro bulbs. Rice plants of variety IR64 were cultivated hydroponically in a controlled environment chamber as described by Hubbart et al (2012) Conditions were 30°C (continuous), photoperiod 12h, with light provided by metal halide bulbs supplemented by incandescent bulbs. Irradiance at plant height was 700 $\Box$mol m^{-2} s^{-1}. Sampling took place at the leaf 11 stage.

Obtaining an initial point cloud

The reconstruction algorithm described in this paper uses an initial point cloud estimate as a basis for the growth of plant surfaces in three dimensions. Numerous software- and hardware-based techniques exist to obtain point clouds of objects. While we have chosen to make use of a software-based technique, patch-based multi-view stereo (PMVS) (Furukawa, 2010), in principle other approaches such as laser scanners or LIDAR systems (Omasa, 2007) could also work effectively in its place.

PMVS reconstructs a three-dimensional point cloud model of a plant and scene, based on multiple two dimensional input images. A requirement of this algorithm is that the intrinsic (focal length, etc) and extrinsic (3D position and orientation) camera parameters be known. In the case of static-camera capture systems, such as that used in RootReader3D (Clark, 2011), calibration (the process of obtaining camera parameters for all images) can be performed once, and the estimated parameters used until the system is reconfigured. Tools exist that can perform automatic calibration of moving camera systems, and our intention is that this software be usable with a single DSLR camera, with images captured manually by a user. We use the VisualSFM (Wu, 2011) system to perform automatic camera calibration. SIFT (Lowe, 1999) features are used to find corresponding points between pairs of images, which are in turn used to reconstruct the 3D positions of each camera relative to all others, and relative to the model being reconstructed. The output is a series of camera models, each associated with a single input image, and the point cloud representation of the scene reconstructed by PMVS.

VisualSFM and PMVS were created to address the general image-based modelling problem and represent the current state of the art in this area. PMVS has been shown to provide high-quality descriptions of simpler, convex objects and, though plants' complexity poses a challenge, provides a sound basis for the development of plant modelling techniques.

An initial surface estimate

Though each point can reasonably be expected to lie on some surface, the point cloud representation produced by PMVS contains no explicit description of those surfaces. If the final plant model is to be used e.g. with a ray tracing system (Song et al., 2013. Zhu et al., 2004), or if details of individual leaves are required, a surface-based description must be constructed. Methods for the reconstruction of a surface mesh from a point cloud exist (Carr, 2001. Kazhdan, 2006). Most, however, construct a single surface describing the entire point cloud. Plants contain large numbers of separate surfaces that must be considered separately; common algorithms such as Poisson surface reconstruction (Kazhdan, 2006) cannot be used over plant canopies, where they attempt to fit a single surface over the plant, they will inevitably

oversimplify the complex structure. In this work we begin by producing an initial surface estimate, before refining this estimate until the plant model is complete.

The initial surface consists of a number of compact planar sections that represent small areas of plant material. Previous work in plant modelling has used planar sections to represent leaves, such as the artificially generated plant models used by Song (2013). Over larger plants, or within a plant canopy, large planes represent an over-simplification of the underlying leaf shape. Smaller surface patches increase accuracy, and allow neighbouring sections with different orientations to characterise the curved nature of the leaves.

In order to establish the required size of the fitted planes, and the location and orientation of each patch, the point cloud is first segmented into small clusters of points (Fig. 1A). We use a radially bounded nearest neighbour strategy (Klasing, 2008) to achieve this, but extend this method to limit the potential size of each cluster. Points are grouped with their nearest neighbours, as defined by a pre-set distance. This distance is dependent on the size and resolution of the model being captured. However, as the PMVS algorithm and laser scanning devices usually output points with a consistent density, the distance parameter can be set once and then remain unchanged between experiments. Here, clusters were limited to a maximum size of 120 points, reducing this number will increase the number of planar sections fitted to the data, increasing accuracy at the cost of decreased algorithmic efficiency. In our experiments, a limit of 120 points represents a significant partitioning of the point cloud, typically producing sections of leaf approximately 2-5cm^2. Crucially, the limited size of the clusters ensures that points on neighbouring leaves, between which a surface should not be formed, are rarely clustered together.

A least-squares orthogonal regression plane is fitted to each cluster of points using singular value decomposition. This best fit plane minimises the orthogonal distance to each point, and is described by a centre point, and two vectors representing the surface orientation and rotation (Fig. 1B).

We identify the co-ordinate system used by our plant model as "world" co-ordinates, representing the 3D geometry of the scene. Each point is associated with the cluster of points to which it has been assigned, and can be projected onto the best-fit plane for that cluster. Once projected, we say that the point resides in 2D "planar" co-ordinates (Fig. 1C). While the orthographic projection between world and planar co-ordinates is in some sense reversible, it should be noted that depth information (the distance to the plane) is purposefully discarded when a point is expressed in planar co-ordinates. This has the effect of flattening the points in each cluster to lie on their best fit plane, reducing noise in individual points, and reducing the search for an optimal patch boundary to two dimensions. It is important, however, to consider the planar and world co-ordinate systems as essentially different views of the same information. As such, point and mesh surfaces generated on a cluster plane will have an associated world position that can be output as a final 3D model.

An initial surface patch description is constructed by calculating the α-shape (Edelsbrunner, 1983) of each set of 2D points, expressed in planar co-ordinates. An α-shape is a subset of the commonly used Delaunay triangulation, but contains restrictions on which edges and faces are preserved. Specifically, an edge between two points is preserved if there exists a generalised disk of radius 1/α, on which those two points lie, that contains the entire point set. In practice, the value α can be increased in size to increase the level of detail in the boundary of the triangulation, by removing larger edges. The α value can be tuned for a given data set, to preserve the shape of the boundary of each reconstructed point set. Figure 1D,E shows the Delaunay Triangulation for the example cluster, and an associated α-shape.

Shape optimisation using level sets

The set of α-shapes computed over each cluster forms an initial estimate of the location and shape of the leaf surface. The limitations of the initial stereo reconstruction on plant datasets means that in many instances this estimate will be inaccurate or incomplete, and will require further optimisation to adequately reflect the true shape of the plant. Missing leaf surfaces should be reconstructed, and overlapping shapes should be optimised to meet at a single boundary. Many methods, such as active contours (Kass, 1988), parameterise the boundary of a shape before attempting this optimisation – fitting a curve which is then manipulated to best describe the leaf. Such approaches are ill-suited to the complex boundary conditions that might be produced by α-shapes. Consider Figure 2A, in which three example clusters have been segmented in close proximity and been described by their α-shape boundaries. The red cluster contains a hole, but this hole should be preserved in order to prevent the yellow cluster from being obstructed. The red cluster also contains two distinct segments, this occurs when all the links between the two sections are longer than α, and is difficult to model with a traditional contour method. This is particularly true where we can expect these two regions to merge as the optimisation process continues, as is likely here.

These problems can be solved using the level set method (Osher, 1988. Sethian, 1996). The distinct regions and hole in the red cluster are preserved, and each region can reshape independently of the others.

The boundary of the α-shape computed for a given cluster is used to initialise a level set distance function. The level set method defines a 3D function φ that intersects the cluster plane (Figure2B-D). We represent this function as a signed distance function, initialised such that negative values lie within our α-shape boundary, and positive values occur outside. Thus, the boundary itself is defined as the set of all points in φ that intersect the cluster plane, given as:

$$\Gamma = \{(x, y)|\varphi(x, y) = 0\}$$

Our goal is to optimise the points in Γ, such that the boundary represents an accurate approximation of the shape of the leaf on which the cluster is located. To achieve this, during each step of the level set process the height of the distance function at any point (x, y) is altered based on a speed function. A negative speed function causes φ to move downwards, causing the boundary Γ to move outwards at that point. Conversely, a positive speed function raises φ, shrinking the boundary. The speed function can be altered based on both global and local parameters, thus different regions of φ can move upward, others downward. The result is a boundary that grows or shrinks as necessary to fit the underlying data, defined by the speed function, and the shape of the boundary Γ is optimised. The change in φ, based on a speed function v is defined as:

$$\frac{\delta\varphi}{\delta t} = -v \cdot |\Delta\varphi|$$

where $\Delta\varphi$ is the gradient of the level set function at a given point. We define our speed function as:

$$v = v_{curve} + v_{image} + v_{inter}$$

v_{curve} is a measure of the local curvature of the boundary, calculated using a central finite difference approximation, as in (Sethian, 1996). The curvature term encourages the boundary of the level set equation to remain smooth. Convex or concave regions will be given negative or positive curvature terms respectively, and will generally become flatter over time.

The image term, v_{image}, references colour information in the input images to ascertain whether the projection of the planar surface lies over regions with a high likelihood of depicting leaf material. This is achieved by using the existing camera geometry to project the positions of the level set function into the camera images. The matrices required to perform this calculation can be derived from the known camera geometry that was used to obtain an initial point cloud reconstruction of the data.

By performing multiple consecutive projections, we are able to examine the pixel information related to a given α-shape from the point of view of any camera. Each planar position can be projected into one or more input images, allowing a combination of colour information from multiple images to be used to calculate v_{image}. However, while it is mathematically possible to project a given planar position into any input image, it may not be useful to do so in all cases. Some images will present an obscured view of a given leaf surface, others may be viewed from a near-perpendicular angle, resulting in an image projection of a region that is very small, and less useful than other camera views. Rather than combining information from multiple views, we choose instead to pick one so-called reference view from which to obtain colour information. We choose a reference image that represents a calculated "best view" of a planar surface in world co-ordinates. This calculation is based on the number of projected, occluded and occluding pixels observed in a cluster when viewed from each camera.

Image selection begins by projecting each cluster into each camera view. Only the interior of each α-shape is projected, using a scan-line rasterisation algorithm to find all pixels in the corresponding image into which that α-shape will project. Attached to each projected position is a *z*-depth value. This records the distance that the projected point lies from the camera's image plane, and can be used to sort clusters that project onto the same location. Projections with the lowest *z* value lie in front of those with higher *z* values. Those with higher *z* values can be marked as occluded, and represent leaf material that cannot be seen in that image. This information is used to calculate the viewable percentage of every cluster, as well as the percentage of each cluster that is occluded, or occluding others relative to each camera view.

It is desirable to select camera views that contain as little interference between clusters as possible. For each image, and each cluster, we calculate $\mathcal{V}^{clear}$, a normalised measure of the number of pixels into which a cluster projects that are not also shared by other clusters, $\mathcal{V}^{occluded}$, representing the percentage of occluded pixels belonging to that cluster, and $\mathcal{V}^{occluding}$, representing the percentage of pixels within a certain cluster that occlude those from other clusters. The clear pixel count also provides a measure of the angle of incidence between a cluster and the camera plane. If a cluster is seen at an angle from a given camera, this camera will likely contain less projected pixels than one that sees a "front on" view of that leaf surface.

For each cluster, the combination of normalised clear pixel count, occlusion and occluding percentages can be used to sort images in terms of view quality:

$$view\ quality = \mathcal{V}^{clear}\left(1 - \mathcal{V}^{occluded}\right)\left(1 - \mathcal{V}^{occluding}\right)$$

A reference image I_R is chosen for each cluster that maximises this view quality measure.

When referencing pixel values in the image I_R, we use a normalised green value (calculated below) to measure the likelihood of leaf material existing at that location. Normalised green is robust to changes in illumination, which are frequent within a canopy where shadows are cast from other leaves.

Computed from the red, green and blue colour channel information, the normalised green value of a pixel is calculated as:

$$normalised\ green = \frac{green}{red + green + blue}$$

As long as care is taken to choose a suitable background during image capture, we can assume that normalised green values will be higher in pixels containing leaf material, and lower in pixels containing background. Where lighting conditions remain consistent over an image set we can also assume that the distribution of normalised green values is the same over each image. However, between different image sets we cannot assume that the properties of the normalised green values are consistent. The distribution of normalised green values, and thus the expected normalised green values of leaf material, must be ascertained before normalised green can be used to contribute to the v_{image} term.

In a dataset where the only strong normalised green response originates from plant material, and the background intensity is fairly consistent, we can assume that the histogram of normalised green values for all pixels over all images will form a bimodal distribution, shown by the dashed line in Figure 3. If we restrict sampling to only those pixels that are projected into by the α-shapes, we would expect the frequency of background pixels to reduce dramatically, as shown by the solid line in Figure 3. Rosin's (2001) unimodal thresholding approach is ideally suited to analysing such a histogram, and is capable of finding an appropriate threshold level below the foreground peak. Using this threshold to separate the two distributions, the mean and standard deviation of the foreground peak are calculated and used to produce a speed function centred around the calculated threshold t, with a spread based on the standard deviation of the peak. More formally, given a normalised green value $\mathcal{N}$:

$$v_{image} = \begin{cases} \max\left(-1, \dfrac{\mathcal{N} - t}{2\sigma}\right), & \mathcal{N} < t \\[2mm] \min\left(+1, \dfrac{\mathcal{N} + t}{2\sigma}\right), & \mathcal{N} \geq t \end{cases}$$

where t is the threshold calculated using Rosin's method, and σ is the standard deviation of the normalised green peak. Essentially the value of v_{image} is 0 where $\mathcal{N} = t$, then decreases to -1 at $\mathcal{N} = t - 2\sigma$, or increases to +1 at $\mathcal{N} = t + 2\sigma$. A spread of 2σ was chosen as a value that characterises the width of the foreground peak.

The final component of the speed function, v_{inter}, works to reshape each surface based on the location and shape of nearby clusters. As neighbouring clusters may lie on surfaces with different orientations, calculation of their 3D intersections is challenging. Indeed, it may be that two nearby clusters that could be considered as overlapping, because their projections into a given image overlap, do not intersect in three dimensions. To avoid this issue we project each planar position into I_R, and examine the interactions in the 2D camera co-ordinate system. Essentially, clusters are considered as overlapping if their projections intersect when viewed from the reference image I_R.

The function v_{inter} is calculated such that each cluster projected into any given point is penalised if it is occluded by another cluster. We define the function as:

$$v_{inter} = \begin{cases} p - v_{image}, & occluded\ in\ I_R \\ 0, & not\ occluded\ in\ I_R \end{cases}$$

where p is a small negative value such that the level set boundary Γ shrinks at this location. Note that the subtraction of v_{image} results in the image component being cancelled out where clusters are occluded. Where a pixel is obstructed in the reference view, the image information should be ignored.

The complete speed function is used to update each discrete position on the level set function φ. This process must be repeated until each cluster boundary has reshaped to adequately fit the underlying image data. The speed function is will slow significantly as we approach this optimal shape. Where a level set boundary no longer moves with respect to the reference image (does not alter the number of projected pixels), we mark this cluster as complete and discontinue level set iterations. Any level sets that do not slow significantly will continue until a maximum time is elapsed, a parameter that can be set by the user. We typically use a value of 200 iterations as a compromise between computational efficiency and offering each level set adequate time to reshape. In many cases clusters will halt naturally before this maximum time has elapsed.

Remeshing the Level Set Functions

Once the level set operation has terminated, clusters' shapes may bear little resemblance to their original α-shape description. Each surface patch must therefore be re-triangulated in order to provide the mesh information required for a complete plant model. The α-shape method used above is less suitable for this task, as the parameter α may have a noticeable effect on the resulting boundary shape: the level set function now has a known boundary that was not available during the original surface estimation. This can be used to drive a more accurate meshing approach that will preserve the boundary contour.

We use constrained Delaunay triangulation for this task, based on the algorithms outlined in (Shewchuk, 2002). A constrained triangulation will account for complex boundary shape when producing a mesh from a series of points, however it will not over-simplify the boundary by fitting surfaces across concave sections, and can consider holes in the surface if required. For each cluster we sample from the boundary of the level set in order to obtain a series of points that travel clockwise around the shape. A constrained triangulation is computed from these points, a process that will automatically generate additional points, where required, within the shape itself. An example mesh generated from a single level set function is shown in Figure 4.

As each point in the new triangulation exists in planar co-ordinates, they can be easily back-projected into world co-ordinates to be output in mesh format. Our software outputs the completed mesh in the standard PLY format, which is readable in all commonly used software packages and can be imported into modelling tools.

Reconstruction of Example Rice and Wheat Datasets

In this section we present results obtained when applying our reconstruction approach to multiple views of single plants. Verification of our approach is achieved using a novel artificial dataset, in which an *in silico* model rice plant is rendered from multiple viewpoints to generate artificial colour images that are then treated in the same way as a real-world image set. This approach allows the reconstructed plant to be directly compared to the artificial target object, a difficult problem if no such ground truth were to exist. The proposed image-based modelling method is also used to create 3D representations of a number of real plants.

We have tested our reconstruction methods on datasets obtained from rice and wheat plants. Images were captured using a Canon 650D DSLR camera with a 35mm lens, at 8 megapixel resolution. The number, and nature of the images were left to the user to decide given the

subject in question, though we recommend more than 30 images surrounding the subject for a single plant. No special consideration was given to the environment in which the plants were imaged, the rice dataset was captured in an indoor environment, the wheat in a glass house. These environments provide complex backgrounds, which raise additional challenges, but the plants can still be reconstructed using our methods. It is likely that a permanent installation with a more strict protocol for image capture would result in more consistent point cloud reconstruction between datasets, readers are encouraged to explore this option if using our methods over extended periods. An overview of each dataset is given in Table 1.

Figure 5 shows the result of applying our reconstruction approach to the rice and wheat datasets. The reconstructions are coloured based on the normal orientation of each surface, the models have not been textured, to avoid concealing imperfections in the output mesh. Quantitative evaluation of the effectiveness of any 3D shoot reconstruction is challenging due to a lack of ground truth models for comparison. Here we offer a qualitative evaluation of the benefits and shortcomings of our approach, followed by a quantitative evaluation using the virtual rice dataset.

The initial surface estimate, obtained by calculating an α-shape over each cluster, will naturally reproduce any flaws present in the PMVS point cloud (Fig. 6A,C). Most notable are the lack of point information in areas of poor texture, and noise perpendicular to the leaf surface, where depth has not be adequately resolved. These issues can be caused by the heavy self-occlusion observed in larger plants, but are often caused in even simple datasets by a lack of image features in the centre of leaves.

Depth noise is significantly reduced by the use of best-fit planes over small clusters, as all points are projected onto a single surface. However, the boundary of each surface is a function of the parameters used to create the α-shape, and the quality of the underlying data. As such, we can expect the α-shape boundaries to be a poor representation of the true leaf shape. With this in mind, we would characterise a successful reconstruction as one that significantly improves upon the initial surface estimate, through the optimisation of the surface boundaries.

Consider Figure 6A,C depicting initial surface estimates from the rice and wheat plants in Figure 5. In Figure 6A, the initial surface contains a great deal of overlapping patches, something that if not rectified would produce errors when quantified or used in modelling. Figure 6B shows our level set approach to have removed this overlap, and smoothed each surface boundary such that the original leaf surface is recovered. This result is representative of the success of our approach across all of the leaves in this dataset.

The wheat dataset generally contains thicker leaves, and their lack of surface patterning often results in only sparse point data in the centre of each leaf. This is shown in Figure 6C, where large missing sections of leaf must be recovered. Figure 6D shows our results on this section, in which the cluster boundaries have been optimised and again form a more continuous surface. There is still some overlap in the clusters on the left of the image, but this is caused by the angle from which the image was rendered, and relates to the relative orientations of neighbouring clusters.

The clusters towards the left of Figure 6D are orientated at slightly disjoint angles, due to noise in the original point cloud reconstruction. This makes optimisation of the inter-cluster boundaries challenging, as the intersection of these boundaries depends not only on their orientation, but also on the position from which they are being viewed (the reference view I_R from the previous section). This is an important characteristic of our reconstruction algorithm in its current form: The boundaries of neighbouring patches will be reshaped relative to the reference camera view I_R for each cluster. It is possible that gaps may be observable between

surfaces when viewed at angles very dissimilar to the reference view. In reality, for clusters with very similar orientations these gaps will be negligible. However, as the main focus of our reconstruction work to date has been optimising the boundary speed functions of level sets, we have yet to address this problem. We anticipate that further work on smoothing the normal orientations of neighbouring clusters or merging neighbouring clusters into a single curved leaf model will solve this issue: this will be a focus of upcoming research.

Quantification of Accuracy Using *in silico* Image Capture

To verify the accuracy of our reconstruction approach, an additional dataset was created based on the plant used in the rice dataset. The rice plant was first manually modelled using the point cloud created by PMVS and 3D graphics software (Topogun, Blender). This is a time consuming and subjective process, and should not be viewed as a suitable alternative to automatic reconstruction. However, it is possible to produce an easily quantifiable ground truth model that can be used as a target for automated reconstruction. We textured and coloured the virtual plant in order to emulate the original plant leaves. Finally, we rendered 40 distinct camera views of the same model (Fig. 7A), simulating an image capture system moving around a static plant. The resulting dataset can then be reconstructed in the same manner as real-world data, but we retain the ability to compare the reconstruction with the original virtual plant, in particular keeping the same co-ordinate system. The original model, and our reconstruction can be seen in Figure 7B,C.

A variety of geometric measures were calculated to quantify the differences between the original ground truth model, and the model output by our reconstruction. In each case the measurements were obtained in world co-ordinates, hence we have shown the relative sizes of these measurements rather than absolute values. It is straightforward to convert arbitrary world units into real-world units such as metres through the use of a calibration target of known size. We use a textured disc of known diameter, placed below the plant we wish to calibrate. The world co-ordinate diameter of this disc in the reconstructed point cloud provides a scaling factor to convert back into real-world units. For the datasets in Figure 5, the dimensions of the conical flask in the rice dataset and the pot in the wheat dataset are known, and can be used to convert into real-world co-ordinates if required.

As seen in Figure 8, the summary measurements of width, depth, height and convex hull volume for the ground truth model and our reconstruction are in good agreement. This shows that our approach does not significantly alter the overall size and shape of the original model. Area measurements were also calculated from both 2D projected viewpoints and an overall surface area measure for each model. The projected sizes were calculated as a normalised, resolution independent pixel count using an orthographic camera model. In all cases the area of the reconstructed model is slightly higher than that of the original ground truth. In the worst case, the corner-projected view shows a 10% increase in size in the reconstruction over the original model, and the total surface area was increased 13% from the original model.

We believe that this increase in surface area is a side effect of the resolution of the input images, and the planar surface approach we have used. Consider a small section of leaf surface that, as in Figure 8, appears 13% larger than we would expect given the underlying plant. With a typical leaf in our dataset being around 20 pixels in width, this represents a boundary cluster moving outwards less than a single pixel, a resolution our level set approach does not currently resolve. One immediate solution to this is to increase the size of the input images, however this would add significant computational overhead.

As we represent our curved leaf surface with a series of planar sections, it is also likely that a small increase in surface area will be observed because these sections are an approximation

to the true surface. In 2D, a series of lines approximating the outer edge of a curve will always be longer than the underlying curve, and this principle extends to our 3D reconstruction. We can reduce the impact of this effect by using smaller cluster sizes when segmenting the point cloud.

While approximating a curved surface with planar sections will inevitably reduce accuracy by a small amount, small sections are easily managed, and the algorithmic complexity of the fitting problem is reduced. Our approach is therefore able to reconstruct more complex surfaces than many existing algorithms applied to this type of data, and crucially is generally applicable to plants of any size or shape. Future work will look to combine this general reconstruction approach with a more specific plant model for certain species, allowing us to obtain more accurate models of larger surfaces, without completely sacrificing generality.

Software Availability

The software associated with this publications is open source, distributed under a Berkeley Software Distribution license. It will be distributed on SourceForge (http://sourceforge.net). A link to the SourceForge distribution page is available at www.cpib.ac.uk. The software is written in C# using the .NET framework, so is currently available only for the Windows operating system.

Conclusions

The recovery of accurate 3D models of plants from colour images is challenging. A single plant constitutes a crowded scene in the sense of Furukawa (2010) and the construction of accurate 3D models of objects of this level of complexity is an active research topic. Images of plants exhibit high degrees of occlusion, with the occlusion relations between leaves varying from image to image. To complicate matters further, individual leaves are difficult to identify: most of the leaves on a given plant have similar colour and texture properties. Rather than address these issues in a single process that transforms a set of images into a three-dimensional model via feature correspondence or silhouette analysis, the approach presented here develops each leaf segment individually, automatically selecting an image likely to contain the necessary information. The proposed method reduces the effect of occlusion by choosing an image with a clear view of the target surface, and addresses the similarity problem by performing detailed analysis of the colours present in that image.

The mesh representation produced provides a detailed model of the surface of the viewed plant that can be used both in modelling tasks and as a route to shoot phenotyping. It should be stressed, though, that the surface description output by the proposed technique comprises a large set of distinct planar patches, rather than larger, curved surfaces describing whole leaves. The level set method re-sizes and re-shapes each patch to maximise its consistency with neighbouring patches and the selected image, and as such the reconstructed patches provide an accurate approximation of the leaf surfaces.

Looking to the future, both field phenotyping and canopy-scale modelling will require 3D models of plant communities and canopies. The major additional challenge as the number of plants is increased is the greater incidence of occlusion between leaf surfaces. Our reconstruction algorithm operates on a "best view" reference image, chosen separately for each patch. It is therefore robust to occlusion, as heavily obscured viewpoints are discarded. However, it is still the case that leaves that cannot be viewed clearly from at least one camera are likely to be poorly reconstructed. This makes the image acquisition process particularly important.

Literature Cited

Alarcon VJ, Sassenrath GF. 2011. Modelling cotton (Gossypium spp.) leaves and canopy using computer aided geometric design (CAGD). *Ecological Modelling*. 222.12:1951-1963.

Alenya G, Dellen B, Torras C. 2011. 3D modelling of leaves from color and ToF data for robotized plant measuring. *IEEE International Conference on Robotics and Automation (ICRA)*.

Blender Foundation. 2013 Blender, v2.69. *www.blender.org*

Carr JC, Beatson RK, Cherrie JB, Mitchell TJ, Fright WR, McCallum BC, Evans TR. 2001. Reconstruction and representation of 3D objects with radial basis functions. *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM.

Clark RT, MacCurdy RB, Jung JK, Shaff JE, McCouch, SR, Aneshansley DJ, Kochian LV. 2011. Three-dimensional root phenotyping with a novel imaging and software platform. *Plant Physiology*. 156.2: 455-465.

Edelsbrunner H, Kirkpatrick, DG, Seidel R. 1983. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*. 29 (4): 551–559

Furukawa Y, Ponce J. 2010. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 32.8: 1362-1376.

Houle D, Govindaraju DR, Omholt S. 2010. Phenomics: the next challenge. *Nature Reviews Genetics* 11.12: 855-866.

Hubbart S, Ajigboye OO, Horton P, Murchie EH. 2012. The photoprotective protein PsbS exerts control over CO_2 assimilation rate in fluctuating light in rice. *Plant Journal*. 71(3):402–412.

Kass M, Witkin A, Terzopoulos D. 1988. Snakes: Active contour models. *International Journal of Computer Vision*. 1.4: 321-331.

Kazhdan M, Bolitho M, Hoppe H. 2006. Poisson surface reconstruction. *Proceedings of the fourth Eurographics symposium on Geometry processing*.

Klasing K, Wollherr D, Buss M. 2008. A clustering method for efficient segmentation of 3D laser data. *IEEE International Conference on Robotics and Automation (ICRA)*.

Kumar P, Cai J, Miklavcic S. 2012. High-throughput 3D modelling of plants for phenotypic analysis. *Proceedings of the 27th Conference on Image and Vision Computing New Zealand*.

Long SP, Zhu XG, Naidu SL, Ort DR. 2006. Can improvement in photosynthesis increase crop yields? *Plant Cell Environment*. 29.3315-330.

Lowe, DG. 1999. Object recognition from local scale-invariant features. *The proceedings of the seventh IEEE international conference on Computer Vision (ICCV)*.

Murchie EH, Reynolds M. 2012. Crop cadiation capture and use efficiency. *In Encyclopaedia of Sustainability Science and Technology*. Springer Science and Business Media.

Omasa K, Hosoi F, Konishi A. 2007. 3D lidar imaging for detecting and understanding plant responses and canopy structure. *Journal of Experimental Botany*. 58.4: 881-898.

Osher S, Sethian JA. 1988. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of computational physics.* 79.1: 12-49.

Quan L, Tan P, Zeng G, Yuan L, Wang J, Kang SB. 2006. Image-based plant modeling. *In ACM Transactions on Graphics (TOG)*. vol. 25, no. 3, pp. 599-604. ACM.

Reynolds R, Foulkes J, Furbank R, Griffiths S, King J, Murchie EH, Parry M, Slafer G. 2012. Achieving yield gains in wheat. *Plant Cell and Environment*. 35:10, 1799–1823.

Rosin PL. 2001. Unimodal thresholding. *Pattern recognition*. 34.11: 2083-2096.

Sethian JA. 1999. Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science. Vol. 3. *Cambridge university press*.

Shewchuk JR. 2002. Delaunay Refinement Algorithms for Triangular Mesh Generation, *Computational Geometry: Theory and Applications*. 22(1-3):21-74.

Song Q, Zhang G, Zhu X. 2013. Optimal crop canopy architecture to maximise canopy photosynthetic CO_2 uptake under elevated CO_2 – a theoretical study using a mechanistic model of canopy photosynthesis. **Functional Plant Biology**. 40, 108–124.

SC Pixelmachine SRL. 2012. Topogun, v2.0. *www.topogun.com*

Watanabe T, Hanan JS, Room PM, Hasegawa T, Nakagawa H, Takahashi W. 2005. Rice morphogenesis and plant architecture: measurement, specification and the reconstruction of structural development by 3D architectural modelling. *Annals of botany*. 95.7: 1131-1143.

White JW, Andrade-Sanchez P, Gore MA, Bronson KF, Coffelt TA, Conley MM, Feldmann KA. 2012. Field-based phenomics for plant genetics research. *Field Crops Research*. 133: 101-112.

Wu C. 2011. VisualSFM: A visual structure from motion system. *http://ccwu.me/vsfm/*

Zhong X, Peng S, Sheehy JE, Visperas RM, Liu H. 2002. Relationship between tillering and leaf area index: quantifying critical leaf area index for tillering in rice. *Journal of Agricultural Science*. 138:269–279.

Zhu XG, Ort DR, Whitmarsh J, Long SP. 2004. The slow reversibility of photosystem II thermal energy dissipation on transfer from high to low light may cause large losses in carbon gain by crop canopies: a theoretical analysis. *Journal of Experimental Botany*. 55, 1167–1175.
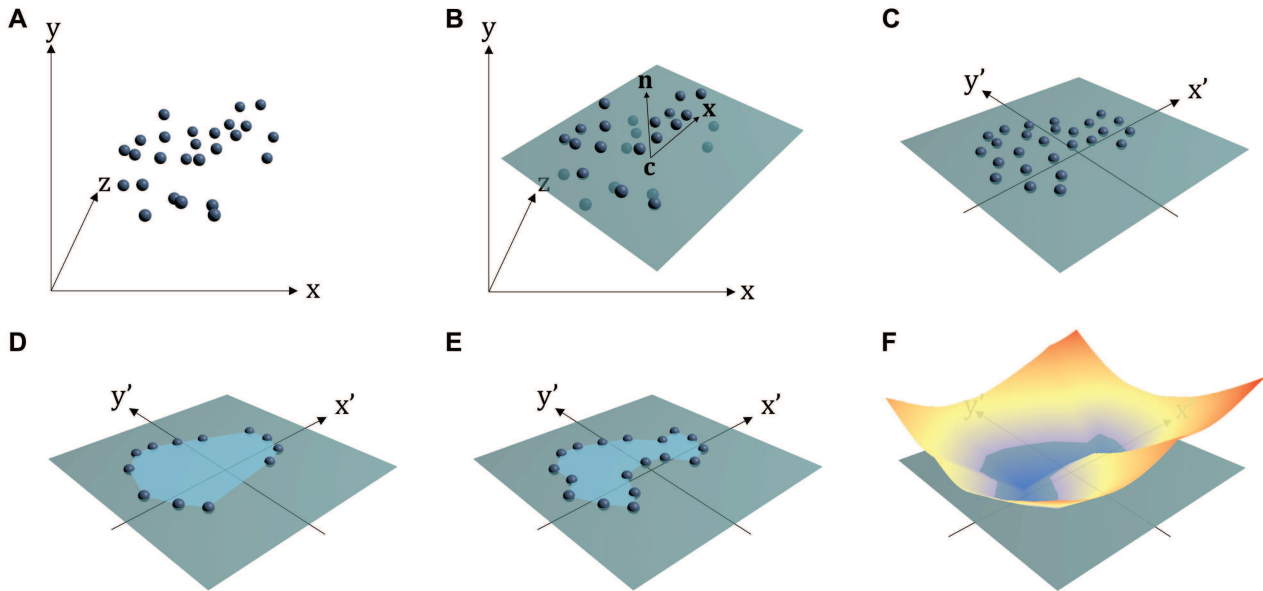
Figure Legends

Figure 1: The process of obtaining an initial surface estimate based on a cluster of points, and initialising a level set method to optimise the surface boundary. (A) A small cluster of points in 3D world co-ordinates, obtained through clustering the input point cloud. (B) An orthogonal regression plane is fitted through the points, **c** represents the centre point of the plane, **n** and **x** represent the orientation and rotation of the plane. (C) Points are orthogonally projected onto the plane surface, in doing so the co-ordinate system is changed to 2D planar co-ordinates. The transformation is such that the x-axis in planar co-ordinates coincides with the vector **x**. (D) The boundary of the Delaunay triangulation of the 2D points in (C). (E) The boundary of an α -shape computed over the same point set. (F) A 3D level set function is initialised such that the intersection with the plane at $z = 0$ resembles the boundary of the α-shape computed in (E).

Figure 2: Level sets allow the complex boundary conditions exhibited within each leaf surface to be modelled. (A) Three α-shapes in close proximity that require shape optimisation. Active contours are ill-suited to this task, due to the disconnected nature of the red region, and the hole containing a separate cluster. (B-D) Individual level set functions intersecting x-y plane at z=0. Using three separate level set functions allows the complex topology of these α-shapes to be preserved, and allows for efficient shape optimisation.

Figure 3: Histogram showing the normalised green distribution over all images in an input set. The solid line includes only locations in which the initial α-shape regions are projected, the dashed line represents all pixels. The dashed line shows a distribution containing a larger amount of background pixels. The reduced frequency of the background in the surface-only distribution can be exploited using Rosin's unimodel thresholding method.

Figure 4: Constrained Delaunay Triangulation of an example leaf surface. (A) The level set φ. Input into the constrained triangulation is a list of points obtained by regularly sampling from the boundary Γ. (B) A constrained Delaunay triangulation preserves the contour of the boundary, and generates additional points inside the surface to complete the mesh.

Figure 5: Output of our reconstruction approach. (A) An image of a wheat plant from the Wheat dataset, a multi-view dataset captured in a glass house. (B) A point cloud representation of the Wheat dataset output by the PMVS software. (C) A reconstructed surface mesh of the complete wheat plant produced by our algorithm. (D) An image of a rice plant from the Rice dataset, a multi-view dataset captured in an office environment. (E) A point cloud representation of the Rice dataset output by the PMVS software. (F) A reconstructed surface mesh of the complete rice plant produced by our algorithm. Both completed meshes have been coloured based on surface orientation for clarity.

Figure 6: Boundary optimisation using level sets. (a) An initial surface estimate on the rice dataset. (b) The reconstructed rice surface after 200 iterations using our level set approach. (c) An initial surface estimate on the wheat dataset. (d) The reconstructed wheat surface after 200 iterations using our level set approach. Regions showing the complex structure of the surfaces have been added above each mesh for clarity.

Figure 7: (A) The model rice plant is rendered from 40 different viewpoints, moving around the plant and from above. (B) The model rice plant, coloured based on the surface normal orientation. (C) The reconstruction of this dataset using our level set approach.

Figure 8: Relative measures of the similarity between the model rice plant, and the reconstructed model obtained by our method. Measures relating to the original virtual plant are shown in black, those of the reconstructed model are shown in grey. (A) Normalised

projected area, measured from four viewpoints. Each image was rendered using an orthographic projection to prevent distortion. This view mimics the parallel light rays produced in our ray tracing system. (B) Geometric measures of the relative sizes of both models. (C) A graph showing the change in model density as a function of the vertical distance through the plant, from the base of the models to the top. The strong peaks represent near-horizontal leaves that cause an abundance of plant material at very specific depths.

Tables

Dataset Name	Image Count	Description
Rice	36	Images of a rice plant (var. IR64) at leaf 10 stage held in a cylindrical flask. Images were captured from a variety of angles surrounding the plant, and from above.
Wheat	64	Images of a wheat plant at growth stage 37 held statically in a growth pot. Images were taken in concentric of increasing height around the plant, at a higher frequency than that of the rice dataset.
Virtual Rice	40	A model plant was manually created based on the Rice dataset. This model was used to generate 40 artificial images from a variety of viewpoints surrounding and above the plant. Unlike real-world examples, the ground truth of this model is known, and can be compared with the output of the reconstruction process.

Table 1: Descriptions of the single plant datasets used to test our reconstruction approach. The Rice, Wheat and Maize datasets were captured manually using a single camera. The virtual rice dataset was generated using 3D modelling software based on a template rice plant created manually. In all cases the input into our reconstruction software was the calibrated camera co-ordinates generated by VisualSFM, the set of input images, and an initial point cloud generated by PMVS.

Figure 1: The process of obtaining an initial surface estimate based on a cluster of points, and initialising a level set method to optimise the surface boundary. (A) A small cluster of points in 3D world co-ordinates, obtained through clustering the input point cloud. (B) An orthogonal regression plane is fitted through the points, **c** represents the centre point of the plane, **n** and **x** represent the orientation and rotation of the plane. (C) Points are orthogonally projected onto the plane surface, in doing so the co-ordinate system is changed to 2D planar co-ordinates. The points are rotated around the normal such that the x'-axis in planar co-ordinates lies along **x**. (D) The boundary of the Delaunay triangulation of the 2D points in (C). (E) The boundary of an α-shape computed over the same point set. (F) A 3D level set function is initialised such that the intersection with the plane resembles the boundary of the α-shape computed in (E).
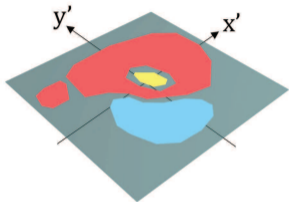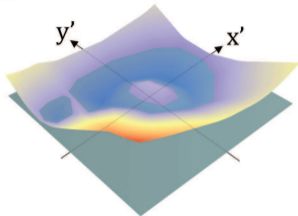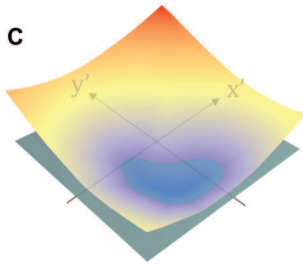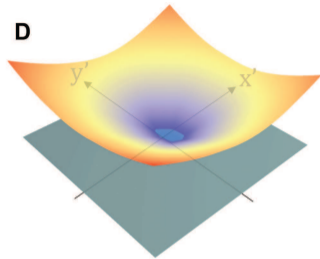
Figure 2: Level sets allow the complex boundary conditions exhibited within each leaf surface to be modelled. (A) Three α-shapes in close proximity that require shape optimisation. Active contours are ill-suited to this task, due to the disconnected nature of the red region, and the hole containing a separate cluster. (B-D) Individual level set functions intersecting x-y plane. Using three separate level set functions allows the complex topology of these α-shapes to be preserved, and allows for efficient shape optimisation.

Figure 3: Histogram showing the normalised green distribution over all images in an input set. The solid line includes only locations in which the initial α-shape regions are projected, the dashed line represents all pixels. The dashed line shows a distribution containing a larger amount of background pixels. The reduced frequency of the background in the surface-only distribution can be exploited using Rosin's unimodel thresholding method, which will select a position below the foreground peak, indicated with the arrow.
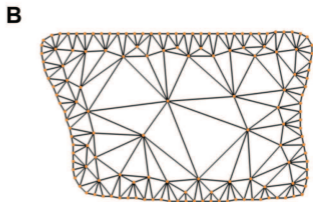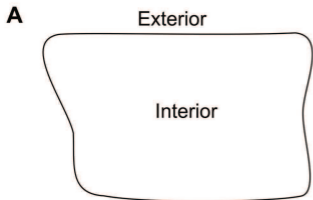
Figure 4: Constrained Delaunay Triangulation of an example leaf surface. (A) The level set φ. Input into the constrained triangulation is a list of points obtained by regularly sampling from the boundary of the level set. (B) A constrained Delaunay triangulation preserves the contour of the boundary, and generates additional points inside the surface to complete the mesh.
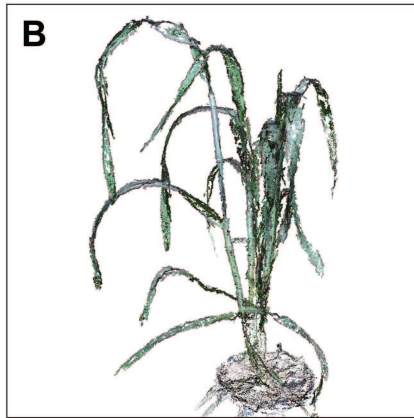
Figure 5: Output of our reconstruction approach. (A) An image of a wheat plant from the Wheat dataset, a multi-view dataset captured in a glass house. (B) A point cloud representation of the Wheat dataset output by the PMVS software. (C) A reconstructed surface mesh of the complete wheat plant produced by our algorithm. (D) An image of a rice plant from the Rice dataset, a multi-view dataset captured in an office environment. (E) A point cloud representation of the Rice dataset output by the PMVS software. (F) A reconstructed surface mesh of the complete rice plant produced by our algorithm. Both completed meshes have been coloured based on surface orientation for clarity.
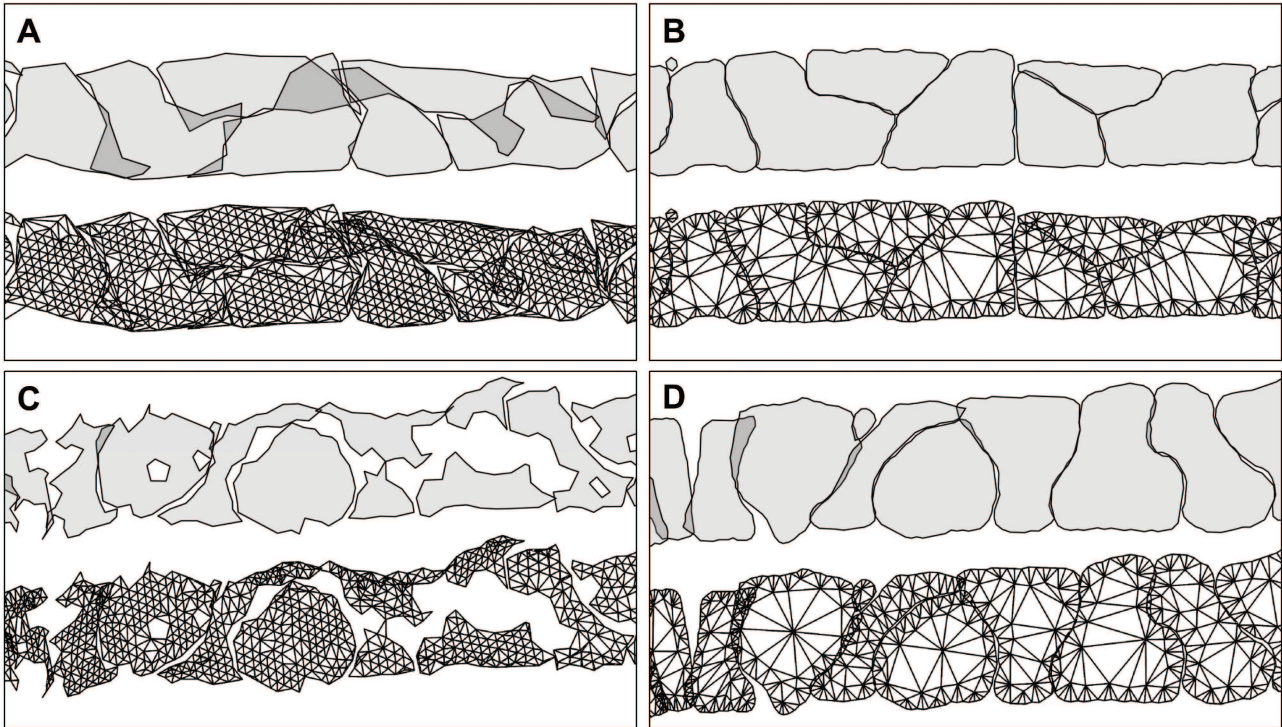
Figure 6: Boundary optimisation using level sets. (A) An initial surface estimate on the rice dataset. (B) The reconstructed rice surface after 200 iterations using our level set approach. (C) An initial surface estimate on the wheat dataset. (D) The reconstructed wheat surface after 200 iterations using our level set approach. Regions showing the complex structure of the surfaces have been added above each mesh for clarity.
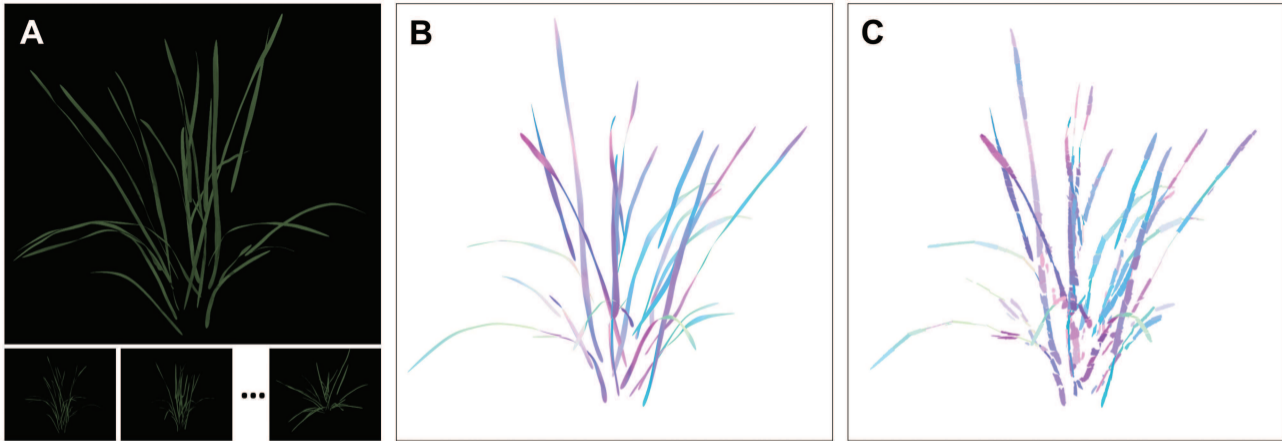
Figure 7: (A) The model rice plant is rendered from 40 different viewpoints, moving around the plant and from above. (B) The model rice plant, coloured based on the surface normal orientation. (C) The reconstruction of this dataset using our level set approach.
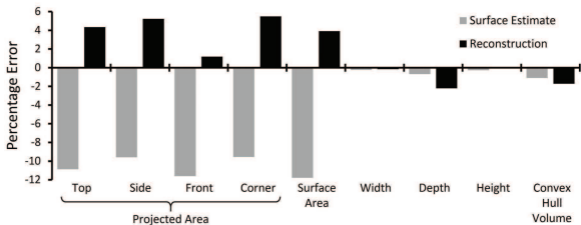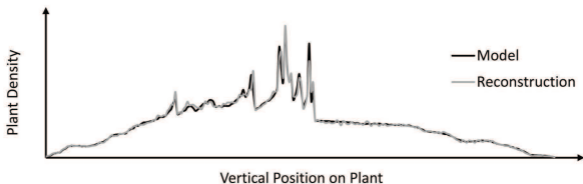
Figure 8: Quantification of the accuracy of our approach on the virtual plant dataset. (A) Percentage error of various measurements between the virtual model, and both the α-shape surface estimate and optimised reconstruction. Lower percentage error indicates that a measurement is closer to the original model. Projected area was measured from four viewpoints, with an orthographic projection mimicking the parallel light rays produced in our ray tracing system. (B) A graph showing the change in model density as a function of the vertical distance through the plant, from the base of the models to the top. The strong peaks represent near-horizontal leaves that cause an abundance of plant material at very specific depths.