3D Plant Modeling: Localization, Mapping and Segmentation for Plant Phenotyping Using a Single Hand-held Camera

Thiago Teixeira Santos(✉), Luciano Vieira Koenigkan, Jayme Garcia Arnal Barbedo, and Gustavo Costa Rodrigues

Embrapa Agricultural Informatics Campinas, Campinas, SP 13083-886, Brazil
{thiago.santos,luciano.vieira,jayme.barbedo,gustavo.rodrigues}@embrapa.br

Abstract. Functional-structural modeling and high-throughput phenomics demand tools for 3D measurements of plants. In this work, structure from motion is employed to estimate the position of a hand-held camera, moving around plants, and to recover a sparse 3D point cloud sampling the plants' surfaces. Multiple-view stereo is employed to extend the sparse model to a dense 3D point cloud. The model is automatically segmented by spectral clustering, properly separating the plant's leaves whose surfaces are estimated by fitting trimmed B-splines to their 3D points. These models are accurate snapshots for the aerial part of the plants at the image acquisition moment and allow the measurement of different features of the specimen phenotype. Such state-of-the-art computer vision techniques are able to produce accurate 3D models for plants using data from a single free moving camera, properly handling occlusions and diversity in size and structure for specimens presenting sparse canopies. A data set formed by the input images and the resulting camera poses and 3D points clouds is available, including data for sunflower and soybean specimens.

Keywords: Plant phenotyping · SLAM, Structure from motion · Segmentation

1 Introduction

In recent years, high-throughput plant phenotyping earned momentum, focusing mainly on non-invasive image-based techniques [17,21,39,40]. Several of these phenotyping efforts rely on some sort of 3D digitizing for the characterization of shoots or roots [12], a useful snapshot of the plant macroscopic state at a precise moment. If the 3D model presents an appropriate sampling of the plant surface at a convenient resolution, different measurements can be performed, such as leaf area, leaf angle and plant topology. Even biomass could be estimated

Electronic supplementary material The online version of this chapter (doi:10.1007/978-3-319-16220-1_18) contains supplementary material, which is available to authorized users.

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part IV, LNCS 8928, pp. 247–263, 2015.
DOI: 10.1007/978-3-319-16220-1_18

using regression methods. Although phenotypes cannot be fully characterized, 3D models are a multipurpose representation: a batch of models produced for leaf area measurements, for example, could be reused in a further study interested in leaf orientation or plant morphology, considering that germplasm and experimental setting data is also available.

A 3D model is also useful for automation. If a robotic device has to interact with a plant, for sensing, pruning or tissue sampling, knowledge about the 3D structure of the plant and the relative pose of the device is essential [2]. For the automation of sensing methodologies such as thermal imaging or chlorophyll fluorescence, pose information can meet requirements regarding sensor orientation relative to the leaf surface.

A system employing 3D modeling for plant phenotyping has to address at least two steps. In the *localization and mapping* step, the sensors' poses must be defined or recovered and the scene 3D model has to be computed. Further, in the *understanding* step, the plant must be detected, its parts segmented and classified, and measurements must be performed.

The present work approaches the first step employing a hand-held, free moving camera. Plant phenotyping systems described in literature use a set of few fixed cameras and a motorized turning table under the specimen to capture images from different poses [34,48]. Those *fixed-camera-moving-plants* approaches are suitable for automation, but present issues regarding occlusion. Alternatively, a *fixed-plant-moving-camera* approach can address occlusions for plants of different species, sizes and development stages. It is also convenient for low-cost phenotyping solutions, since digital cameras are affordable and ubiquitous. A structure from motion (SfM) framework is employed: visual local features in video frames are detected in the plant surface and in surrounding objects, matched across frames and used to recover the camera pose and the 3D location using projective geometry and robust estimation techniques [20]. Once the camera's poses are defined, multiple-view stereo can be used to get a sampling for the plant's surface with greater resolution.

This work also presents alternatives for two parts of the understanding step: segmentation of plants' parts and surface estimation. After segmentation, the surface of each part is estimated by fitting trimmed B-splines to the segment's 3D points. For measurements such as leaf area, a proper surface representation is more adequate than a set of irregularly-spaced points.

2 Related Work

The first attempts to perform 3D digitizing for plants employed mechanical devices [25], sonic digitizers [47] or magnetic trackers [38]. These contact methods were important on the development of functional-structural models for plant development [19], but are unable to scale for high-throughput phenotyping. Current phenotyping initiatives generally rely on contactless methods such as laser scanning using LiDAR devices [9,10], Time-of-Flight (ToF) cameras [1,50] or stereo vision [6,34,44,48].

Kaminuma *et al.* [24] employed a laser range finder for building three-dimentional models for *Arabidopsis thaliana*. Leaves and petioles were represented as polygonal meshes. The meshes were used to quantitatively determine two morphological attributes: the direction of the leaf blade and leaf epinasty, in order to characterize two different ecotypes. The setting was able to produce a good sampling for surfaces: the distance and the sample size allowed a resolution of 0.045 mm per pixel, producing a dense 3D points cloud. However, different plants presenting larger dimensions, or even *Arabidopsis* specimens in more advanced stages of development, could produce a sparse set of points.

The work of Ivanov *et al.* [23] is possibly the first work in the literature using stereo vision to reconstruct the 3D surface of a plant for measurement and analysis. The authors estimated the position, orientation and area of maize leaves (*Zea mays L.*). Unfortunately, the difficulties imposed by the equipment available at the time (digital photography was not yet widespread) undermined the experiments. The segmentation of the leaves and determination of correspondences between images were performed manually, using photographic enlargements. Despite the limitations, this work was the forerunner of more recent systems, which employ more up-to-date advances in computing performance, digital imaging and computer vision. Biskup *et al.* [6] developed a stereo vision system based on two digital cameras to create three-dimensional models of soybean plants foliage, aiming to analyze the angle of inclination of leaves and their movement throughout the day. Given the importance of movement for the experiment, the system was able to process up to three images per second, recovering the needed 3D information for slope computation.

Depth data from ToF cameras can be fused to RGB data from common cameras to produce 3D reconstructions for leaves, despite the low resolution of ToF devices. Song *et al.* [50] fused data from a stereo RGB camera pair (480×1280 pixels) and a ToF camera (64×48 pixels) using the graph-cut energy minimization approach [7]. The stereo pair provided a higher resolution while the ToF information addressed disparity failures caused by occlusion and matching on textureless areas. Alenyà *et al.* [1] employed a simpler fusion approach: the 3D points from ToF data were transformed to the coordinates of the RGB camera reference frame, and then projected to the camera image plane - color points not having a 3D counterpart were discarded. Although simple, this approach was able to produce good results because (i) the employed ToF camera presents a greater resolution (200×200) and (ii) a robotic arm could move the cameras to new positions and acquire more data, under the *sensing-for-action* method proposed by the authors.

Under the stereo vision framework, a camera from the stereo pair can be replaced by a projector. The same ray intersection principle can be applied if the projector position and intrinsics are known and the pixel correspondences can be found. Bellasio *et al.* [5] used a calibrated camera-projector pair and a coded-light method to find pixel correspondences, recovering the 3D surface of pepper plant leaves. Chéné *et al.* [8] employed a RGB-D camera (a Microsoft Kinect device) to segment leaves and estimate their orientation and inclination. These

RGB-D devices are camera-projector triangulation-based systems assembled in a single device, also using coded-light to infer pixel correspondences [16].

Recently, multiple-view stereo (MVS) has been applied on plant digitizing to address the occlusion problems found in 3D plant reconstruction. Paproki *et al.* [35] employed the 3D S.O.M. software [3] to create 3D models for cotton plants and then estimate stem heights, leaf widths and leaf lengths. The cotton specimens were placed in a rotating tray presenting a calibration pattern used by the 3D S.O.M. software to estimate the camera relative position at each frame.

Santos and Oliveira [44] employed the structure from motion (SfM) framework [20] to recover the camera calibration data. SfM extends the stereo vision framework incorporating a step that simultaneously look for the best estimation for camera positions *and* 3D point locations. This process, called *bundle-adjustment* [20,51], is a global maximum-likelihood estimation process that evaluates all the multiple-views at the same time, minimizing the re-projection error for each view. Instead of using a calibration pattern, the authors employed a local feature detection and matching approach combined to bundle-adjustment to recover the camera parameters, as proposed by Snavely *et al.* [49] in the context of city architecture 3D modeling. This camera data and the initial sparse point cloud was employed as input for the patch-based MVS (PMVS) algorithm [18], which produced dense point clouds that represent the surface of leaves and internodes in experiments with basil, *Ixora* [44] and mint [43].

Sirault *et al.* [48] also reported 3D reconstruction results using the PMVS system. In their work, the calibration data came from a setting using fixed cameras and potted plants placed in a very precise turntable (2.5 million counts per revolution). Their digitizing platform, PlantScan^{TM}, is a chamber equipped with three RGB CCD cameras, a NIR camera and two LiDAR laser scanners, plus two termobolometer sensors that collect thermal information. Their stereo reconstruction fuses PMVS, voxel coloring [26] and LiDAR data using registration algorithms (RANSAC [13] and ICP [41]). The authors report their platform is able to digitize plants from a few centimeters up to two meters height and up to a meter thick, enabling the phenotyping of a broad range of different species.

Fixed camera settings can be cursed by occlusion problems. In contrast, a free-moving camera has more flexibility on getting images from different poses, adapting the acquisition to different plant architectures and sizes. A calibration artifact [3] could be used to provide the landmarks needed to camera localization. However, considering the diversity of plant architectures and sizes, the use of such artifacts would impose hard constraints in the camera acquisition, since they must be visible in every image. The present work relies on visual local features observed on the plant surface (and on objects nearby) to provide the landmarks for camera localization. A chessboard pattern is used just to normalize the scale and orientation of the final 3D model, and there is no need to observe such pattern in every image.

3 Method

The methodology starts with a computer-aided image acquisition step that helps the user on getting a set of images suitable for 3D reconstruction (Section 3.1). The images (video frames) and the local features acquired are employed in a localization and mapping step performed under the structure from motion framework, determining the camera parameters for each image and an initial 3D model for the scene (Section 3.2). A multiple-view stereo step takes the images and camera parameters for each pose to produce a dense 3D point cloud sampling the plant's surface (Section 3.3). Clustering is used to segment the point cloud, isolating the individual leaves, and then B-spline interpolation is employed to approximate the surface of each leaf (Section 3.4).

3.1 Computer-Aided Image Acquisition

A phenotyping setting employing a free moving camera is able to get poses that allow a better handling of occlusions and variations on plant size and morphology. The work of Alenyà *et al.* [1] illustrates this idea, showing how proper path-planning for a robotic arm can move the cameras to positions where occlusions can be easily solved (in their case, using ToF cameras and robot's positioning control).

In previous attempts on 3D plant modeling by SfM [43,44], hundreds of photographies were taken and used as input. This image capturing procedure was executed by human operators that spent up to 30 minutes on this acquisition step. Furthermore, the operator could discover some images out of focus or lacking local features, or even situations in which the wide-baseline between images produced poor feature matching, making the scene reconstruction impossible.

The aim of the proposed system for computer-aided image acquisition is to help the user to quickly produce an image set suitable for SfM-based 3D reconstruction using a video camera. This image set should present feature-rich images and good feature correspondence between them. Video streams produce thousands of frames in a few minutes, so the method automatically selects frames, ensuring a reasonable number of feature correspondences between the selected ones. There is a short-baseline/wide-baseline trade-off: too short baselines between successive frames would produce an enormous image set. However, too wide baselines would produce poor feature matching due to plant auto-similarity. Algorithm F describes the procedure.

Algorithm F (*Frame Selection*). This procedure produces a sequence of *keyframes* $\mathbf{K} = \langle K_1, K_2, \ldots K_n \rangle$ such that there is a proper local features matching between any pair of successive frames K_i and K_{i+1}. A feature matching is considered proper if (i) at least 10% of the found local features in K_{i+1} were properly matched to features in K_i and (ii) the estimated fundamental matrix `F` can properly map 70% of the matched features (inliers rate).

F1. [*Initialize.*] At the initialization, grab the first keyframe, K_1. Consider to select a frame presenting a large number of features.

F2. [*Frame process.*] Grab an input frame F. Detect local features using SURF [4] and compute their descriptors. Consider to use a GPU based implementation of SURF to ensure processing at the frame-rate.
F3. [*Feature matching to the last keyframe*]. Match the feature descriptors found in frame F to the ones in the last keyframe K_n, using the Lowe's criteria (see [28] and Section 3.2). Set r_{match} to the ratio of successfully matched features.
F4. [*Compute fundamental matrix.*] Compute the fundamental matrix $\mathtt{F}$ using RANSAC for robustness (see [20]). Set r_{in} to the ratio of inliers, i.e. the ratio of the points consistent to $\mathtt{F}$[1].
F5. [*Frame selection.*] If $r_{\text{match}} > 0.1$ and $r_{\text{in}} > 0.7$, then $F_{\text{last}} \leftarrow F$. Otherwise, if $F_{\text{last}} \neq \text{NIL}$, push F_{last} into $\mathbf{K}$ and set $F_{\text{last}} \leftarrow \text{NIL}$. Return to step F2. ▌

F_{last} is the last frame (i) presenting a good matching to the last keyframe (the test in step F5) and (ii) that is not yet a keyframe ($F_{\text{last}} \notin \mathbf{K}$). If the current frame F does not present a proper matching to the last keyframe and $F_{\text{last}} = \text{NIL}$, then the system is in *lost state*. The system will present the last keyframe to the user, who have to move the camera to new positions until a new grabbed frame F presents a good matching, turning the system back to a *tracking state*. The user can drop keyframes from $\mathbf{K}$, in an attempt to help on seeking a new frame that put the system in the tracking state again. Figure 1 (c) shows the acquisition user interface.

3.2 Visual Odometry

Visual odometry is the process of estimating the ego-motion of an agent [45], in this case a single hand-held camera. As the present work is concerned about 3D reconstruction of plants, global map consistency regarding the camera poses and the object is essential, so such an odometry is a simultaneous localization and mapping (SLAM) problem. The local features act as landmarks for the localization step, recovering the relative pose between two frames. The mapping step is performed registering the features in the current map, thus defining the absolute pose of the current frame. Global consistency can be enforced by bundle adjustment [20,51].

Windowed Feature Matching. As the local features are employed as landmarks, their positions in different frames have to be found. In the SfM framework [20], the feature correspondences between two frames K and K' allow a robust estimation of the fundamental matrix $\mathtt{F}$. Furthermore, $\mathtt{F}$ is used on the computation of the relative pose, in the form of two projection matrices $\mathtt{P}$ and $\mathtt{P}'$.

[1] $\mathtt{F}$ maps a point $\mathbf{x}$ in frame K to an *epipolar line* $\mathbf{l}' = \mathtt{F}\mathbf{x}$ in frame K'. A pair $\mathbf{x}, \mathbf{x}'$ is consistent to $\mathtt{F}$ iif the distance between $\mathbf{x}'$ and the line $\mathbf{l}$ is under a defined threshold. See [20] for details.

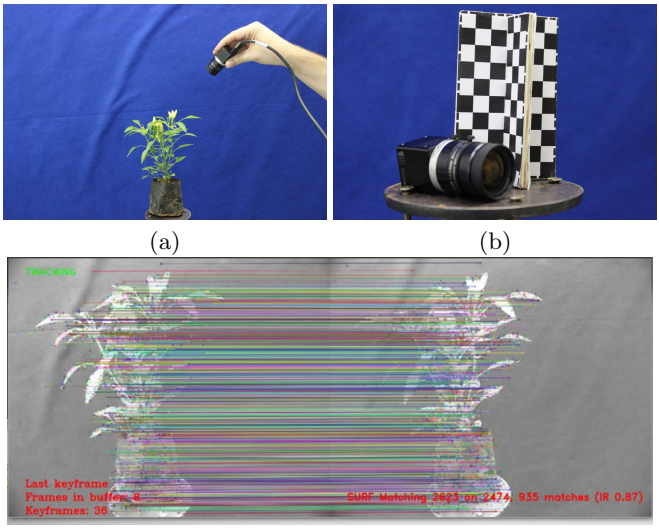

Fig. 1. Computer-aided image acquisition for structure from motion. (a) User freely moves a camera around the specimen, getting images from different poses to solve occlusions. (b) The employed machine-vision camera and a pattern used to recover the scale, essential to further metrology. (c) The acquisition system interface aids the user.

The local feature matching is performed using the stored SURF local features and their descriptors computed in the acquisition step, as presented in Algorithm F, step F2. The employed matching criteria is the one presented by Lowe [28]. For a feature descriptor f in K, its two nearest neighbors f'_1 and f'_2 are found in K', presenting Euclidean distances d_1 and d_2 respectively. A match (f, f'_1) is declared if $\frac{d_1}{d_2} < 0.6$ and if there is no other feature in K that matches f'_1 in K'.

For fast nearest neighbors searches, a KDTree is employed. Unfortunately, KDTree performance degenerates for large dimensions. Using PCA, the original SURF descriptors, which are 64-element vectors, are transformed into shorter vectors of dimension 16. Such dimensionality reduction did not degenerate the feature matching: just a few matches are lost if compared to the original 64-D descriptor space.

As discussed previously in Section 3.1, auto-similarity in plants can produce spurious matches between features from different frames in the wide-baseline case. Instead of looking for matches for each possible pair of frames, a *windowed matching* is employed: for each frame F_i, the matching is performed relative to frames $\langle F_{i-m}, ..., F_{i-2}, F_{i-1} \rangle$. An example of feature matching between two near frames is shown in Figure 2.

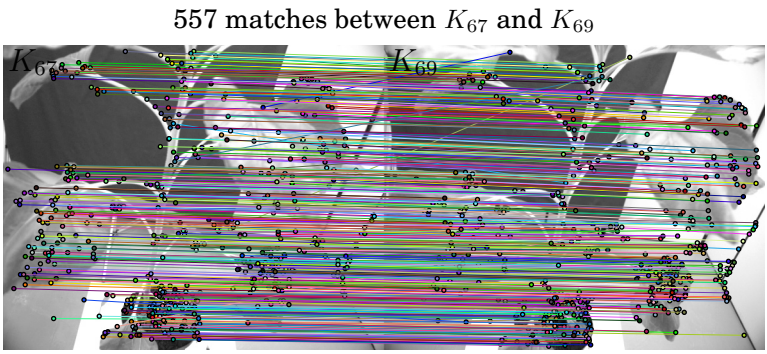

Fig. 2. Local feature matching between two keyframes K_{67} and K_{69} from the keyframes sequence produced for a sunflower specimen using Algorithm F. Local feature matching is performed using Lowe's method [28]. The robust RANSAC procedure can estimate the fundamental matrix `F`, considering wrong matches as outliers.

SfM. Considering the chain formed by the matches, $(f, f'), (f', f''), \ldots$, it is possible to define feature *tracks*. The localization and mapping procedure is performed by the SfM system called *Bundler*, developed by Snavely *et al.* [49]. At each step, Bundler selects the pose that observes the largest number of tracks whose 3D locations have already been estimated. This pose's extrinsic parameters are estimated by the direct linear transform (DLT) described by Hartley and Zisserman [20] under a RANSAC procedure for robustness. New 3D points are added to the map by triangulating the matching features between the new pose and the previous ones. After each pose estimation step, Bundler performs a bundle adjustment optimization, using the SBA package by Lourakis and Argyros [27].

Bundler's standard procedure employs David Lowe's SIFT implementation [28] for feature detection and matching. Matching is performed for *every combination of two frames*, and the intrinsic and extrinsic parameters for each pose are estimated. In this work, this procedure is modified: the windowed matching using SURF features is employed, and the intrinsic parameters are set to fixed values. If a pair of nearby frames (i.e., in a same window) presents less than 128 matches, it will not be considered for extrinsic parameters computation, since it could produce a poor estimation. In relation to the intrinsic parameters, Bundler was designed for SfM using pictures from different consumer cameras. In this work, the same fixed focus camera is moved around the plant, so the intrinsic parameters are the same for every camera pose.

In the experiments, Bundler produced some spurious camera poses. These wrong poses can affect the subsequent multiple-view stereo step, when a dense 3D model is computed. Fortunately, those wrong poses are detached from the right camera path. DBSCAN clustering [11] is employed to group camera locations in tracks. This clustering algorithm is able to group elongated clusters of

points together, making it suitable to identify a long camera track and isolate spurious poses forming short tracks. Spurious camera positions and the 3D points triangulated from them are removed and a final bundle adjustment is performed for the remaining poses and points. The adjusted poses and the set of frames are then used as the input for the multiple-view stereo step.

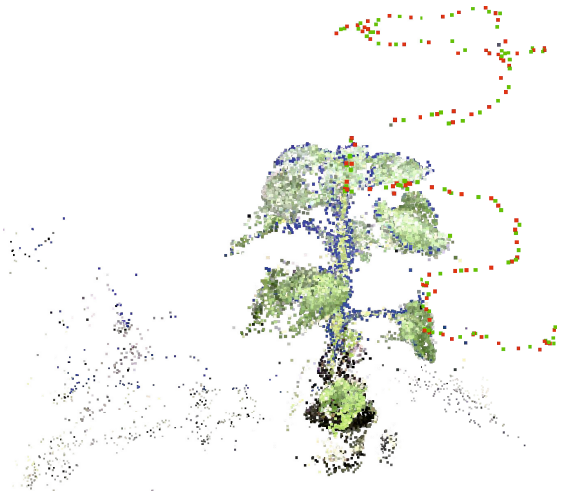

Fig. 3. Structure from motion results for the *sunflower* sequence. The red and light green squares show the camera position for each input frame. The scene mapping is produced incrementally by the triangulation of the local features found, using the recovered pose information. Even in this sparse map, it is possible to see the plant's leaves, internodes and its pot.

3.3 Multiple-View Stereo

Camera poses retrieved in the previous step can be used by a multiple-view stereo (MVS) method to produce a dense 3D model. The method used here is the patch-based MVS (PMVS) proposed by Furukawa and Ponce [18]. This method produces a set of *surfels*, oriented 3D patches that cover a small area of an object surface (the surfel orientation should correspond to the surface normal vector). PMVS is composed by matching and expanding steps that create new patches and estimate their surface orientation while enforcing local photometric consistency, and a filtering step that removes erroneous patches, enforcing visibility consistency (exploring the camera poses and the patches projections). This point cloud (or surfels cloud) is used in the next step for segmentation.

Object Segmentation. No screen or shield was employed to form a standard background behind the plant during image acquisition. Such shields could constrain the user movement around the specimen. Furthermore, features identified

on other objects in the scene actually help the SfM step, providing information from different depths that improves the pose estimation. However, after the 3D reconstruction, it is necessary to identify what elements of the point cloud that correspond to the objects of interest: the plant, its pot and the reference pattern (employed in a further step for scale correction). Again, DBSCAN clustering provides a proper solution: the objects of interest form a dense and possibly elongated cluster in the scene. Because the camera is always pointing to the plant, the objects of interest form the densest and largest cluster in the point cloud.

Scale Correction. To recover a proper scale for measurements and to place the model in a standard orientation, a chessboard-like pattern was inserted in the scene, near the plant. Two nearby frames where the pattern is visible are randomly selected. Eight reference points on the planar pattern are manually marked in each frame (Figure 4) and used to estimate a homography matrix $\mathtt{H}$ [20]. This matrix is able to transform points on the pattern plane in frame K to their corresponding ones in frame K'. The matrix $\mathtt{H}$ is used to find the points $\mathbf{x}'_i$ in K' corresponding to points $\mathbf{x}_i$ on the pattern in image K, ensuring the pairs $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ are on the *same plane* in the scene. This is important because the method is relying on the 3D location of such reference points to properly align the model to a standard coordinate frame in millimeter scale.

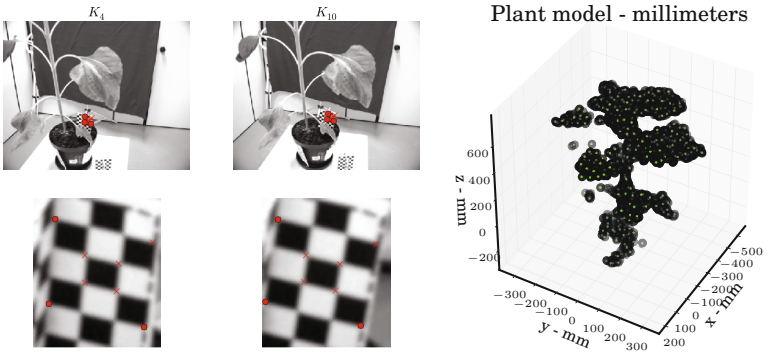

Fig. 4. Changing the model scale and orientation. Eight points on the pattern were marked in frames K_4 and K_{10}. The estimated homography ensures the points are in the same plane in both images. Three points (shown as circles) are used to define a new scale and orientation for the model. After the transformation (right image), measurements in the model can be performed in millimeters.

Fig. 5. Point clouds corresponding to the 3D reconstructions by multiple-view stereo for *sunflower* (left and middle) and *soybean* (right). (a) Examples of input frames, 1280 × 1024 pixels images captured by a machine vision camera (XIMEA model MQ013MG-E2). (b) 3D models: a 65,145 points model for the *sunflower/first take* data (left), a 48,974 points model for the *sunflower/second take* data (middle), and a 167,382 points model for the *soybean* data (right). (c) 3D models (top view). Point clouds are available in the supplementary material set as Stanford Triangle Format files – PLY.

3.4 Plant Segmentation

Segmentation is not just an essential part of the plant model analysis, but it can also be considered the *first* step. Once segmented and classified, each part can be further processed under a specific context. For example, internodes could be modeled by generalized cylinders and leaves by curve fitting.

The point cloud produced by the MVS step is segmented by *spectral clustering* [29,33,46]. Spectral clustering is a suitable option to problems where the centers and spreads are not an adequate description of the cluster. The method uses the top eigenvectors of an affinity matrix. Here, an affinity matrix is built computing, for each point in the cloud, the similarity to its 20 nearest neighbors (based on the Euclidean distance between the points' 3D locations). Examples of the clustering results can be seen in Figure 6 and Figure 7.

Fig. 6. Spectral segmentation for the *sunflower* model. Left: side view. Right: top view.

Surface Estimation. For each leaf point cloud, a surface is estimated using non-uniform rational B-splines (NURBS) [37]. As the leaf surface presents two main orientations, the B-spline surface is initialized using PCA to define such principal directions, adding few control points. In a refinement phase, new control points are added. Further, the surface is trimmed by a B-spline curve that encloses the points as proposed by Flöry and Hofer [14]. The entire procedure is described by Mörwald [30] in the Point Cloud Library documentation. Figure 7 shows some results for *soybean-leaves*.

4 Data and Results

Frame acquisition was performed using a XIMEA machine vision camera (model MQ013MG-E2) equiped with a EV76C560 sensor (resolution of 1280×1024 pixels), and 6 mm fixed focal length lenses (Fujinon model DF6HA-1B). Acquisitions were performed for a sunflower specimen (two takes), forming the *sunflower* dataset, and for a soybean specimen: a take covering the entire plant, the *soybean* dataset, and a second take, closing in a few leaves, the *soybean-leaves* dataset.

During the acquisition step, local feature detection, description and matching were performed using the GPU SURF implementation available in the OpenCV

library (version 2.4). Structure from motion was performed using Bundler, version 0.4 [49] and multiple-view stereo using PMVS, version 2 [18]. DBSCAN and spectral clustering were performed using scikit-learn [36], version 0.14, and NURBS fitting using the Point Cloud Library (PCL) [42], version 1.7. All data (images, recovered camera poses, 3D models and segmentation) is publicly available[2]. The 3D models can be found in the supplementary material, as files in the Stanford Triangle Format (PLY)[3].

Figure 5 shows the 3D reconstructions for *sunflower* and *soybean*. Each plant presents height larger than 50 cm and, for each keyframe, just a part of the plant is visible. This contrasts with the small plant scenario found in previous works [43,44], where the entire plant could be focused and placed in the camera field of view. The visual features detected in the plant surface were sufficient to recover the camera pose and map the entire specimen. Thin structures as the soybean internodes could be properly reconstructed (see supplementary data for a detailed view). The complexity and density of the soybean shoot demonstrates the advantages of a free-moving camera approach.

Figures 6 and 7 show the segmentation results. The graph-cuts performed by the spectral clustering are able to isolate the leaves, being effective for both soybean and sunflower plants. However, spectral clustering needs the desired number of segments be previously defined.

The 167,392 points model in Figure 5 and the 62,677 points model in Figure 7 correspond to the same soybean plant. Approximating the camera to the leaves produced a denser and more detailed 3D reconstruction. This means closer acquisitions could produce denser models for the full plants, however this would make the acquisition step longer and more tiresome for human operators.

Fig. 7. Leaf 3D modeling, segmentation and surface estimation using NURBS for the *soybean-leaves* dataset. Left: a 62,677 points model created by moving the camera near a set of 9 leaves in the sunflower specimen. Middle: segmented model. Right: trimmed NURBS surfaces.

[2] https://www.agropediabrasilis.cnptia.embrapa.br/web/plantscan/datasets

[3] PLY files can be visualized and manipulated using the open-source software Meshlab available at http://meshlab.sourceforge.net.

5 Conclusions

A single hand-held, free-moving camera can be used to create accurate three dimensional models for plants by using state-of-the-art structure from motion and multiple-view stereo techniques. This methodology is suitable for automatized plant phenotyping facilities or small laboratories with restricted resources.

The presented SfM methodology relies on local features detected in the plant surface acting as visual landmarks. Unfortunately, the lack of visual features in some plants can make camera localization unfeasible. Experiments with maize specimens, presenting almost featureless surfaces, were unsuccessful. Structured light approaches, that create "artificial features" using light patterns[4], or ToF cameras [15] would be more appropriated in those cases. For smaller plants, visual markers could be employed [3].

Better feedback could be provided in the acquisition step if real-time reconstruction was performed. Such information would let the user to evaluate model completeness and resolution level, allowing proper camera posing. Real-time reconstruction systems for RGB [32] and RGB-D [22,31] sensors are promising alternatives in this direction.

Systems based on monocular SfM, LiDAR, RGB-D and ToF cameras are providing lots of data and moving the computer vision field to the next step: the analysis of 3D point clouds for automatic scene understanding. This is also true for plant models, because phenotyping pipelines need to make sense of such 3D data for phenomic characterization. Automated segmentation, classification and metrology of each plant part are current research issues.

Acknowledgements. This work was supported by Brazilian Agricultural Research Corporation (Embrapa) under grants 03.11.07.007.00.00 (PlantScan) and 05.12.12.00 1.00.02 (PhenoCorn). We would like to thank Katia de Lima Nechet (Embrapa Environment) for her support providing plant specimens.

References

1. Alenya, G., Dellen, B., Torras, C.: 3D modelling of leaves from color and ToF data for robotized plant measuring. In: 2011 IEEE International Conference on Robotics and Automation, pp. 3408–3414, May 2011
2. Alenya, G., Dellen, B., Foix, S., Torras, C.: Robotized Plant Probing: Leaf Segmentation Utilizing Time-of-Flight Data. IEEE Robotics & Automation Magazine **20**(3), 50–59 (2013)
3. Baumberg, A., Lyons, A., Taylor, R.: 3D S.O.M.A commercial software solution to 3D scanning. Graphical Models **67**(6), 476–495 (2005)
4. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding **110**(3), 346–359 (2008)
5. Bellasio, C., Olejníčková, J., Tesa, R., Sebela, D., Nedbal, L.: Computer reconstruction of plant growth and chlorophyll fluorescence emission in three spatial dimensions. Sensors (Basel, Switzerland) **12**(1), 1052–1071 (2012)

[4] This is the method behind RGB-D devices as the Microsoft Kinect.

6. Biskup, B., Scharr, H., Schurr, U., Rascher, U.: A stereo imaging system for measuring structural parameters of plant canopies. Plant, Cell & Environment **30**(10), 1299–1308 (2007)
7. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(11), 1222–1239 (2001)
8. Chéné, Y., Rousseau, D., Lucidarme, P., Bertheloot, J., Caffier, V., Morel, P., Belin, E., Chapeau-Blondeau, F.: On the use of depth camera for 3D phenotyping of entire plants. Computers and Electronics in Agriculture **82**, 122–127 (2012)
9. Dassot, M., Colin, A., Santenoise, P., Fournier, M., Constant, T.: Terrestrial laser scanning for measuring the solid wood volume, including branches, of adult standing trees in the forest environment. Computers and Electronics in Agriculture **89**, 86–93 (2012)
10. Delagrange, S., Rochon, P.: Reconstruction and analysis of a deciduous sapling using digital photographs or terrestrial-LiDAR technology. Annals of Botany **108**(6), 991–1000 (2011)
11. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD, vol. 96, pp. 226–231 (1996)
12. Fiorani, F., Schurr, U.: Future scenarios for plant phenotyping. Annual Review of Plant Biology **64**, 267–291 (2013)
13. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981)
14. Flöry, S., Hofer, M.: Constrained curve fitting on manifolds. Computer-Aided Design **40**(1), 25–34 (2008)
15. Foix, S., Alenya, G., Torras, C.: Lock-in Time-of-Flight (ToF) Cameras: A Survey. IEEE Sensors Journal **11**(9), 1917–1926 (2011)
16. Freedman, B., Shpunt, A., Machline, M., Arieli, Y.: Depth mapping using projected patterns. US Patent 8,493,496, July 23, 2013
17. Furbank, R.T., Tester, M.: Phenomics-technologies to relieve the phenotyping bottleneck. Trends in Plant Science **16**(12), 635–644 (2011)
18. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(8), 1362–1376 (2010)
19. Godin, C., Costes, E., Sinoquet, H.: A Method for Describing Plant Architecture which Integrates Topology and Geometry. Annals of Botany **84**(3), 343–357 (1999)
20. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, April 2004
21. van der Heijden, G., Song, Y., Horgan, G., Polder, G., Dieleman, A., Bink, M., Palloix, A., van Eeuwijk, F., Glasbey, C.: SPICY: towards automated phenotyping of large pepper plants in the greenhouse. Functional Plant Biology **39**(11), 870 (2012)
22. Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D.: RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments. The International Journal of Robotics Research **31**(5), 647–663 (2012)
23. Ivanov, N., Boissard, P., Chapron, M., Andrieu, B.: Computer stereo plotting for 3-D reconstruction of a maize canopy. Agricultural and Forest Meteorology **75**(1–3), 85–102 (1995)

24. Kaminuma, E., Heida, N., Tsumoto, Y., Yamamoto, N., Goto, N., Okamoto, N., Konagaya, A., Matsui, M., Toyoda, T.: Automatic quantification of morphological traits via three-dimensional measurement of Arabidopsis. The Plant Journal **38**(2), 358–365 (2004)
25. Lang, A.: Leaf orientation of a cotton plant. Agricultural Meteorology **11**(c), 37–51 (1973)
26. Leung, C., Appleton, B., Buckley, M., Sun, C.: Embedded voxel colouring with adaptive threshold selection using globally minimal surfaces. International Journal of Computer Vision **99**(2), 215–231 (2012)
27. Lourakis, M.I.A., Argyros, A.A.: SBA: A software package for generic sparse bundle adjustment. ACM Transactions on Mathematical Software **36**(1), 1–30 (2009)
28. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision **60**(2), 91–110 (2004)
29. Luxburg, U.: A tutorial on spectral clustering. Statistics and Computing **17**(4), 395–416 (2007)
30. Mörwald, T.: Fitting trimmed B-splines to unordered point clouds. http://pointclouds.org/documentation/tutorials/bspline_fitting.php
31. Newcombe, R.a., Davison, A.J., Izadi, S., Kohli, P., Hilliges, O., Shotton, J., Molyneaux, D., Hodges, S., Kim, D., Fitzgibbon, A.: KinectFusion: Real-time dense surface mapping and tracking. In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality, pp. 127–136. IEEE, October 2011
32. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: DTAM: dense tracking and mapping in real-time. In: 2011 International Conference on Computer Vision, pp. 2320–2327, November 2011
33. Ng, A., Jordan, M., Weiss, Y.: On spectral clustering: Analysis and an algorithm. Advances in neural information ... (2002)
34. Paproki, A., Fripp, J., Salvado, O., Sirault, X., Berry, S., Furbank, R.: Automated 3D segmentation and analysis of cotton plants. In: 2011 International Conference on Digital Image Computing: Techniques and Applications, pp. 555–560, December 2011
35. Paproki, A., Sirault, X., Berry, S., Furbank, R., Fripp, J.: A novel mesh processing based technique for 3D plant analysis. BMC Plant Biology **12**, 63 (2012)
36. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
37. Piegl, L., Tiller, W.: The NURBS Book. Monographs in Visual Communication, U.S. Government Printing Office (1997)
38. Rakocevic, M., Sinoquet, H., Christophe, A., Varlet-Grancher, C.: Assessing the Geometric Structure of a White Clover (Trifolium repens L.) Canopy using3-D Digitising. Annals of Botany **86**(3), 519–526 (2000)
39. Rascher, U., Blossfeld, S., Fiorani, F., Jahnke, S., Jansen, M., Kuhn, A.J., Matsubara, S., Märtin, L.L.A., Merchant, A., Metzner, R., Müller-Linow, M., Nagel, K.A., Pieruschka, R., Pinto, F., Schreiber, C.M., Temperton, V.M., Thorpe, M.R., Dusschoten, D.V., Van Volkenburgh, E., Windt, C.W., Schurr, U.: Non-invasive approaches for phenotyping of enhanced performance traits in bean. Functional Plant Biology **38**(12), 968 (2011)

40. Reuzeau, C., Frankard, V., Hatzfeld, Y., Sanz, A., Camp, W.V., Lejeune, P., Wilde, C.D., Lievens, K., de Wolf, J., Vranken, E., Peerbolte, R., Broekaert, W.: Traitmill^{TM}: a functional genomics platform for the phenotypic analysis of cereals. Plant Genetic Resources **4**(01), 20–24 (2006)
41. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: Proceedings Third International Conference on 3-D Digital Imaging and Modeling, pp. 145–152 (2001)
42. Rusu, R.B., Cousins, S.: 3D is here: point cloud library (PCL). In: IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, May 9–13 (2011)
43. Santos, T., Ueda, J.: Automatic 3D plant reconstruction from photographies, segmentation and classification of leaves and internodes using clustering 1. In: Sievänen, R., Nikinmaa, E., Godin, C., Anna Lintunen, P.N. (eds.) Proceedings of the 7th International Conference on Functional-Structural Plant Models, Saariselkä, Finland, pp. 95–97 (2013)
44. Santos, T.T., de Oliveira, A.A.: Image-based 3D digitizing for plant architecture analysis and phenotyping. In: Saúde, A.V., Guimarães, S.J.F. (eds.) Workshop on Industry Applications (WGARI) in SIBGRAPI 2012 (XXV Conference on Graphics, Patterns and Images). Ouro Preto (2012)
45. Scaramuzza, D., Fraundorfer, F.: Visual Odometry [Tutorial]. IEEE Robotics & Automation Magazine **18**(4), 80–92 (2011)
46. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(8), 888–905 (2000)
47. Sinoquet, H., Moulia, B., Bonhomme, R.: Estimating the three-dimensional geometry of a maize crop as an input of radiation models: comparison between three-dimensional digitizing and plant profiles. Agricultural and Forest Meteorology **55**(3–4), 233–249 (1991)
48. Sirault, X., Fripp, J., Paproki, A., Kuffner, P., Nguyen, C., Li, R., Daily, H., Guo, J., Furbank, R.: PlantScan: a three-dimensional phenotyping platform for capturing the structural dynamic of plant development and growth. In: Proceedings of the 7th International Conference on FunctionalStructural Plant Models, Saariselkä, Finland, pp. 45–48 (2013)
49. Snavely, N., Seitz, S., Szeliski, R.: Modeling the World from Internet Photo Collections. International Journal of Computer Vision **80**(2), 189–210 (2008)
50. Song, Y., Glasbey, C.A., van der Heijden, G.W.A.M., Polder, G., Dieleman, J.A.: Combining stereo and time-of-flight images with application to automatic plant phenotyping. In: Heyden, A., Kahl, F. (eds.) SCIA 2011. LNCS, vol. 6688, pp. 467–478. Springer, Heidelberg (2011)
51. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment – a modern synthesis. In: Triggs, B., Zisserman, A., Szeliski, R. (eds.) ICCV-WS 1999. LNCS, vol. 1883, p. 298. Springer, Heidelberg (2000)