

Leaf segmentation from ToF data for robotized plant probing

G. Alenyà, B. Dellen, S. Foix, and C. Torras

Abstract—Supervision of long-lasting extensive botanic experiments is a promising robotic application that some recent technological advances have made feasible. Plant modelling for this application has strong demands, particularly in what concerns 3D information gathering and speed. This paper shows that Time-of-Flight (ToF) cameras achieve a good compromise between both demands. A new method is proposed to segment plant images into their composite surface patches by combining a hierarchical segmentation of the infrared intensity image, provided by the ToF camera, with quadratic surface fitting using ToF depth data. Leaf models are fitted to the segments and used to find candidate leaves for probing. The candidate leaves are ranked, and then the robot-mounted camera moves closer to selected leaves to validate their suitability to be sampled. Some ambiguities arising from leaves overlap or occlusions are cleared up in this way. Suitable leaves are then probed using a special cutting tool also mounted on the robot arm. The work is a proof-of-concept that dense infrared data combined with sparse depth as provided by a ToF camera yields a good enough 3D approximation for automated cutting of leaf discs for experimentation purposes.

I. INTRODUCTION

Recent advances in depth sensors [1], deformable object modelling [2], and autonomous mobile manipulation [3] have considerably widened the scope of robot application. One area that is nowadays gaining attention since it could benefit from all these advances is the monitoring and maintenance of large botanic experimentation fields, e.g., for plant phenotyping. The goal is to determine the best treatments (watering, nutrients, sunlight) to optimize predefined aspects (plant growth, seedling, flowers) and, towards this aim, experiments entailing many repetitive actions need to be conducted [4]. Measurements and samples from leaves must be regularly taken and some pruning may need to be performed [5]. These are tasks for which robots would be very handy, however, difficulties arise from the complex structure and deformable nature of plants, which do not only change appearance through growing, but whose leaves move also on a daily cycle.

In the last twenty years, several robotic systems have been introduced for the automated harvesting of tomatoes, cucumbers, mushrooms, cherries, strawberries, and other fruits (for a review see: [6]), but these systems have not yet reached the stage of commercialization due to the challenges posed by the task. The automated probing of plant leaves is a related, but new research topic in agricultural robotics with many potential applications. For example, probes could be taken from plants automatically in order to detect plant disease or nutritional deficiencies. Treatment of singular plants can then prevent spreading of disease in fields and reduce the

application of chemicals. Another potential application is the fast probing of plants in research laboratories for phenotyping purposes. We expect to face similar challenges as the ones previously encountered with picking robots in agriculture: (i) the recognition and localization of the target, e.g, fruits and leaves, given the varying appearances of plants, and (ii) the probing, grasping, cutting or detachment of parts of the plant under weakly constrained conditions in natural environments [7], [8]. Another major challenge in agricultural robotics is the guidance of motions through crop fields or greenhouses, which is a topic that is not addressed in this work.

The first challenge requires new solutions for the recognition and localization of leaves to be developed. Previously color vision has been used to obtain some relevant plant features, mainly for recognition and classification purposes [9], but when it comes to extracting structural/geometric information for 3D modelling and robot manipulation, the concurrence of a user is required to provide hints on segmentation from multiple views [10]. If a fully automated process is sought, depth information needs to be extracted through stereo [11], structured light [12] or a laser scanner [13]. These techniques have proven adequate for offline modelling, but either require special conditions or are too slow to be used in online robot interaction with plants. Recently, Time-of-Flight (ToF) cameras have been proposed as a good alternative [14], since they provide low-resolution depth images at 25 frames-per-second. This permits quickly acquiring and fusing images from different viewpoints [15], which is very useful since one-shot plant data are often partial or ambiguous. Some works have exploited the best of both technologies by combining ToF data with high-resolution color images to deliver dense depth maps [16], [17].

Concerning robot action, planning and learning algorithms for the manipulation of deformable objects [18] are deemed to play an important role in this context. Planning needs to encompass the motion of the camera as well, since plants are prone to occlusions and merging of close leaves, so that selecting the best next viewpoint may be crucial to disocclude leaves [19], [20] as well as to determine and access suitable probing points.

More precisely, we address the problem of accurately placing a cutting tool on a leaf, in order to acquire sample discs from several plants at different developmental stages, so as to subsequently analyze their relative growth rates [21]. Thus, the emphasis of this work is on *sensing-for-action* methods developed to segment leaves, fit quadratic surfaces to them, determine best candidates for probing, move the cameras to get a closer view, determine a suitable sampling point on the chosen leaf, and finally reach this point with a disc-cutting tool. Intensity-based segmentation is complemented with depth

Authors are with Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain; {galenya, bdellen, sfoix, torras}@iri.upc.edu



Fig. 1. A. WAM arm used in the experiments holding the ToF sensor, a color camera (data not used), and the cutting tool used to extract samples from the leaves. B-C. Typical intensity image and color-coded 3D point cloud acquired with a ToF camera (200×200 PMD CamCube 3.0).

data supplied by a ToF camera to delimit and fit surface patches to the leaves. The ToF camera and the cutting tool are mounted on the robot end-effector (as shown in Fig. 1), so that an egocentric coordinate frame is used for all motions.

II. OVERVIEW OF THE METHOD

The probing of a leaf follows a two-stage approach (see Fig. 2). Initially, the robot arm is moved to a position from which a general view of the plant is obtained. The depth and infrared images acquired with a ToF camera (Section III) are segmented into their composite surfaces as described in Section IV. Leaf-model contours are fitted to the extracted segments, the validity of the fit and the graspability of the leaf are measured, and the segments are ranked (see Section V). A target leaf is selected and the robot moves the camera to a closer, fronto-parallel view of it. The validity of the target and the graspability are re-evaluated (see Section V-A and V-B). If the leaf is considered to be suitable for being sampled based on these criteria, the probing tool is placed onto the leaf following a two step path (see Section V-C). If the target is considered to be non-suitable for probing, another target leaf (from the general view) is selected and the procedure is repeated.

III. 3D IMAGE ACQUISITION

Depth measurements are acquired by a ToF camera (see Fig. 1A-B). This type of sensor has the main advantage of providing registered depth and infrared-intensity images of a scene at a high frame-rate. ToF cameras use the well-known time-of-flight principle to compute depth. The camera emits modulated infra-red light in order to measure the travelling time between the known emitted waves and the ones reflected back over the objects in the scene.

ToF cameras have two main drawbacks: low resolution (e.g. 200×200 pixels for a PMD CamCube 3.0 camera) and noisy depth measurements due to systematic and non-systematic errors [22]. On the one hand, low resolution can be a big problem for large environment applications, but it has not such a negative impact when the camera is used at close ranges as it is our case. On the other hand, noisy depth measurements due to non-systematic errors get amplified by working in such a short range. Mainly the ones due to multiple light

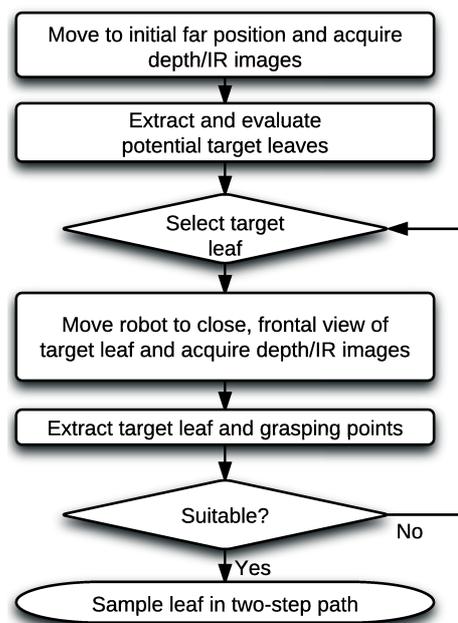


Fig. 2. Flow diagram of the suggested probing procedure (see Section II).

reception and light scattering. Systematic errors get highly reduced by calibration procedures and non-systematic ones can be palliated using filtering techniques [23].

Here we apply two filters to remove undesired wrongly estimated point depths and noise: a jump edge filter and an averaging filter [24]. Sometimes these false measurements are indicative of possible model misinterpretation or object occlusion and, therefore, their detection and 3D localization in the scene may provide valuable information for computing the next-best-view that can help to disambiguate or improve occluded leaf visibility and pose estimation [25].

IV. DEPTH SEGMENTATION

In this section we describe an algorithm for segmenting the sparse and noisy depth data measured by the ToF camera into surface patches in order to extract task relevant image regions, i.e., leaves. We assume that plant leaves are usually represented by a single surface in 3D space. While this assumption may not be generally valid, we however assume that it holds in most cases. Due to the many occlusions present in grown plants and the variability of leaves in terms of size, orientation, and 3D shape, the application of appearance models directly to the image data with the purpose of leaf segmentation would be extremely challenging, also since partial shape models might have to be utilized.

Removing noise and invalid points in the depth data using the jump-edge filter provides a sparse depth map. We segment the data by using the infrared-intensity image of the depth sensor as an auxiliary image. Unlike depth, which is measured using the ToF principle, the corresponding infrared-intensity image provides complete (dense) information with little noise. In comparison with color or respective gray-level images, the infrared intensity images are more amenable to segmentation,

since plant-type characteristic color textures are not present here. The segments are then selected and merged based on the available depth information, which can be sparse.

The algorithm proceeds as follows. First, the infrared-intensity image is segmented with a standard algorithm at different resolutions. Details can be found in [26]. This is necessary since we do not know beforehand at which resolution good regions will appear. Those segments which fit the depth data best, according to a parametric surface model (see Section IV-A), are selected, and a new segmentation is constructed. This procedure has been described in detail in [20], and will thus not be repeated here. From this intermediate segmentation and the respective estimated parametric surfaces, a graph is build, where the nodes represent segments and edges represent the pairwise similarity of the segments surfaces, as described in Section IV-B. Then, to remove remaining over-segmentations present in the intermediate segmentation, a graph-based merging (clustering) procedure is employed which allows us to handle the non-local character of surface properties (see Section IV-C and IV-D). An overview of the algorithm is provided in Figure 3.

The method requires currently about ≈ 28 s to segment an image and to fit surface models using Matlab and non-optimized code.

A. Fitting of quadratic surface models

For modelling the 3D surfaces of image regions, we use a quadratic function, which allows (among others) the modeling of planar, spherical, and cylindrical shapes. Surfaces with more involved curvatures could also be managed within the same approach, but are not required for the application at hand. Moreover, we use quadratic functions that allow computing depth z explicitly for the x - y coordinates in the form of $z = f(x, y)$. This way, surfaces are described by five parameters a, b, c, d , and e , where the depth z can be expressed as a function of x and y through $z = ax^2 + by^2 + cx + dy + e$.

For a given segment s_i we perform a minimization of the mean squared distance

$$E_{i,\text{model}} = 1/N \sum_j (z_j - z_{j,m})^2 \quad (1)$$

of measured depth points $z_{j,m}$ from the estimated model depth $z_j = f_{i,\text{model}}(x_j, y_j)$, where $f_{i,\text{model}}$ is the data-model function and N is the number of measured depth points in the area of segment s_i . The optimization is performed with a Nelder-Mead simplex search algorithm provided in MATLAB.

B. Segment graph

A nearest-neighbor graph is constructed from the image segments. For each image segment, the boundary points are extracted and the local neighborhood within a radius of 1 px of each point is searched for points belonging to other segments that lie within a predefined absolute depth distance. For computing the depth distance, the fitted depth derived for the respective segment point is used, i.e. $|f_i(x_i, y_i) - f_j(x_j, y_j)|$, where i and j denote neighboring pixels belonging to different segments, respectively. Two segments are considered

neighbors if the respective boundary points are less than $d_{3D} = 1$ cm apart. The segments define the nodes V of the segment graph (V, e) . An edge e exists between two segments if they are neighbors according to the condition given above.

C. Segment dissimilarity

We define a dissimilarity measure e_d between two segments s_i and s_j by estimating how well the surface model of segment s_i describes the depth data of segment s_j and vice versa. Let f_i be the surface model of segment s_i , and f_j the surface model of segment s_j . Then, we compute the fitting errors

$$E_{i/j} = 1/n_i \sum_{p \in s_i} [f_j(x, y) - z(x, y)]^2, \quad (2)$$

and

$$E_{j/i} = 1/n_j \sum_{p \in s_j} [f_i(x, y) - z(x, y)]^2, \quad (3)$$

where $z(x, y)$ is the measured depth at (x, y) , $f_j(x, y)$ and $f_i(x, y)$ are the estimated depth value using surface models at (x, y) , and n_i and n_j are the number of points in segment i and j , respectively. Note that the surface parameters have been estimated before, hence no surface fitting has to be performed at this step. Then smaller error is selected, yielding e_d .

D. Graph-based merging of segments

The pairwise dissimilarities between segments are used to sort the graph edges e_{ij} in order of increasing dissimilarity. For this purpose, we define a label l enumerating the edges in ascending order. The total number of edges is n . We further define a merging threshold d_{merge} , which in our case should be chosen in the range between 1 and 5 cm² to be in proportion to the expected range of target fitting errors in the given scenario. The surface models of all graph nodes or segments are also stored in a list, because they may be updated during the procedure.

Then the algorithm proceeds as follows.

- (1) We select the first edge of the ordered list labeled $l = 1$.
- (2) The two segments linked by the edge labeled l are merged if the edge dissimilarity $e_d(l) < d_{\text{merge}}$. In this case, a new region $s_{i \cup j}$ is created and the respective surface model $f_{i \cup j}$ is found. The surface models of region s_i and s_j are replaced by the new surface model $f_{i \cup j}$. A flag is set indicating whether the surface model of a segment has been updated or not. If however $e_d(l) \geq d_{\text{merge}}$, nothing needs to be done.
- (3) We select the next edge of the ordered list labeled $l = l + 1$. If one of the segments linked by the respective edge has been updated previously and thus flagged, the edge dissimilarity between the segments is recomputed using the current surface models.
- (4) Step 2-3 are repeated until $l = n$.

Working consecutively along the ordered list and updating the surface models along the way allows us to avoid testing

for all possible merging combinations, which may easily lead to a combinatorial explosion. This strategy gives preference to merges of segments with large similarity. The method is related to Kruskal’s algorithm for finding the minimum spanning tree of a graph [27], with the main difference that certain graph edges have to be updated after each merge.

V. EXTRACTION OF GRASPING POINTS

We assume that the procedure described above delivers segments that correspond to leaves of the plant. This assumption may not always hold, but it is a good enough working hypothesis as we will demonstrate below.

The goal of this work is to identify and model leaves from ToF data in order to find suitable grasping points and approach vectors for probing. We use the following strategy. First, a target segment is selected from the processed data obtained from a far (general) view of the plant. Using the surface normal and 3D position of the target, we move the robot arm with the mounted ToF camera to a closer position to the target and align the viewing direction of the camera with its surface normal. At this close position, a new image is acquired, which we use to confirm or reselect our target. If a suitable leaf target is found, a grasping point is identified and an approach to the leaf is planned.

For probing a leaf, two main requirements have to be met by the grasping point for the task to be executable:

- (i) The grasping point should lie within the part of the leaf that points away from the stem of the plant. This way, the risk of collisions with the stem and other leaf parts can be reduced. We further want to approach the leaf from the side to maximize the touched leaf area.
- (ii) The grasping point should not be occluded and or obstructed by other leaves (or objects) in the vicinity of the points.

To fulfill requirement (i), a leaf-specific contour needs to be fitted to the leaf segment boundary in order to map leaf-specific grasping points along the segment boundary (see Section V-A). The contour fitting error here gives us a measure of validity of the selected points. The grasping points from (i) are further tested for their graspability using criteria (ii) (see Section V-B). Both the contour fitting error and the graspability measure are important for evaluating whether a planned grasp is executable.

The contour fitting and grasp point identification requires about 2 s for a single segment using Matlab and non-optimized code.

A. Contour fitting for grasping-point identification

We extract the outer 2D boundary C_i of segment i , consisting of a set of points $\{x, y, z\}$. Before C_i can be compared with the model boundary, we need to rotate the boundary in 3D to a predefined orientation, that aligns its surface normal with the z -axis. This way, perspective distortions can be removed at least partly, leading to a point set $\{x, y\}_r$, where we ignore variations in the z -coordinate, since we are only interested in the projection of the leaf boundary onto the x - y plane.

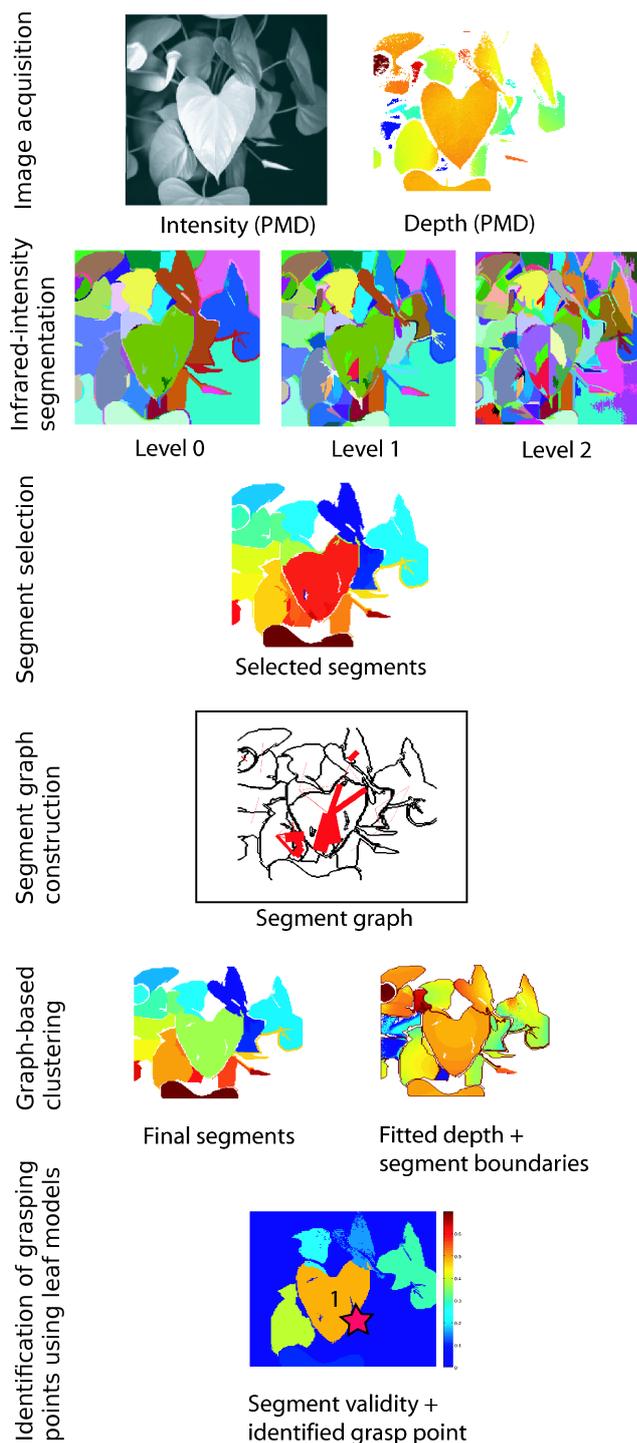


Fig. 3. Schematic of the leaf-extraction algorithm. ToF data (depth and infrared intensity) is acquired and the infrared-intensity image is segmented at different resolutions (level 0-2). Surface models are fitted to the segments and those segments along the segmentation hierarchy that fit the depth data best are selected. From the selected segments a segment graph is constructed and a graph-based segment merging procedure is employed. Final segment contours are fitted to predefined model contours and grasping points are determined.

For each plant type, we have extracted the leaf boundary which is characteristic for the specific plant. We smooth the boundary points with a Gaussian function. The resulting values provide a set of weighted boundary points $\{x, y, w\}_m$, defining our model boundary C_m .

Compared to the model boundary C_m that is characteristic for a specific leaf, C_i might be translated, rotated, or scaled in 2D. These three transformations provide four parameters, i.e., a translation vector (x_t, y_t) , a rotation angle θ , and scaling factor α . Applying these transformations to C_i leads to a transformed set of points $\{x, y, z\}_{\text{trans}}$. The distance of the transformed boundary to a model boundary for given transformation parameters is defined as

$$D(C_i, C_m) = n_i - \sum_{\mathbf{p}_k \in C_m} \sum_{\mathbf{p}_j \in C_i} w_k \delta(x_j - x_k) \delta(y_j - y_k) / n_i + \sum_{\mathbf{p}_k \in C_m} \sum_{\mathbf{p}_j \in C_i} w_k \delta[\delta(x_j - x_k) + \delta(y_j - y_k)] / n_m \quad (4)$$

with $n_m = \sum_{\mathbf{p}_k \in C_m} w_k$ and $n_i = \sum_{\mathbf{p}_j \in C_i} 1$, where $\delta(a) = 1$ if $a = 0$ and zero otherwise, and $\mathbf{p}_k = (x_k, y_k)$ and $\mathbf{p}_j = (x_j, y_j)$. This distance measure decreases the more points of the segment contour are matched to the model contour and increases the more points of the model contour are unmatched.

We find the parameters of the transformations that provide a best match to the model contour by minimizing the distance $D(C_i, C_m)$ using a Nelder-Mead simplex search algorithm provided in Matlab. Once the segment contour has been fitted to the model contour, we can identify grasping points. We assume that predefined grasping points are provided together with the leaf contour model, as illustrated in Figure 4. For each model grasping point, we find the point on the segment contour that has the smallest distance to the model grasping point. Together with the resulting grasping point $\mathbf{x}_g = (x_g, y_g, z_g)$, we also provide the validity measure of the fit.

B. Graspability of identified grasping points

We consider a grasping point (which by definition here lies on the boundary of the segment) to be graspable, if there are no obstructing objects, i.e. other leaves, in its direct vicinity, and if the given boundary is a true leaf boundary, i.e., it is not caused by an occlusion. We define a graspability measure by counting (negatively) the points in a circular area (in 2D) around the grasping point that belong to another segment and are located within a predefined threshold distance d (here, $d = 10$ cm) from the grasping point, or have a depth value z smaller than z_g , yielding

$$g(x_g, y_g, z_g) = - \sum \Theta[\Theta(d - |\mathbf{x} - \mathbf{x}_g|) + \Theta(z_g - z)] \times |1 - \delta[s_l(x_g, y_g) - s_l(x, y)]|, \quad (5)$$

where $\Theta(a) = 1$ if $a > 0$ and zero otherwise, and $s_l(x, y)$ and $s_l(x_g, y_g)$ are the segment labels of points (x, y) and (x_g, y_g) , respectively. The radius of the circular area around the grasping point is chosen equal to d .

C. Intermediate goal position and probing point

To probe the leaf, the probing tool needs to be placed such that the leaf can slide during the approach into the cavity of

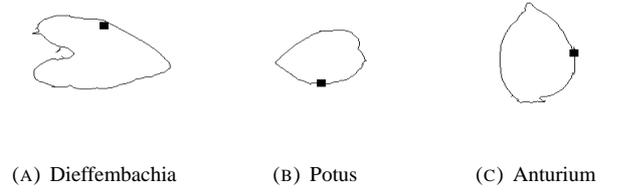


Fig. 4. Model contours for different plants used in our experiments with associated grasping points. Dieffenbachia was used for experiments (A-C), Potus for experiment D, and Anturium for experiment E. Model contours have been extracted from single selected leaves.

the tool, which is only two centimeters wide. For this approach to be successful, the probing tool needs to be aligned with the orientation of the leaf. For this purpose, the average surface normal of the leaf is computed. Furthermore, the probing tool needs first to be placed at an intermediate goal position at a certain distance of the grasping point. We compute the intermediate goal position by first defining an approach vector for the grasp according to $\mathbf{a}_g = \mathbf{x}_g - \mathbf{x}_c$, where \mathbf{x}_c is the center point of the leaf. The approach vector is normalized and used together with the grasping point to compute the intermediate goal position $\mathbf{x}_{\text{goal}} = \mathbf{x}_g + 10\mathbf{a}_g$, at 10 centimeters distance from the edge point towards the outside of the leaf.

We further define a probing point at which the tool should be finally placed $\mathbf{x}_{\text{probing}} = \mathbf{x}_g - 2\mathbf{a}_g$. The probing point is located at 2 centimeters distance from the edge point towards the inside of the leaf.

VI. EXPERIMENTAL SETUP

The experimental setup includes a PMD CamCube Time-of-Flight camera and a PointGrey Flea camera rigidly attached to the last link of a Barrett WAM arm (Fig. 1). The PointGrey Flea camera is however not used in the experiments here. As can be observed, the cameras are displaced from the robot end-effector position to leave room for a cutting tool we have designed to take samples of some selected leaves.

We have opted for a configuration where the cutting tool is outside the field of view of the camera. This implies that, during the robot motion from the close view of the leaf to the placement of the cutting tool, the leaf is not in the camera field of view, and the motion is then performed in open loop. Implicitly we are assuming that the leaf will not move and that the robot has enough precision along this small motion.

The robot and plant initial relative configuration assures that the plant's region of interest is reachable by the robot's cutting tool. In the same way that plant position is guaranteed to be inside the field of view of the camera's initial pose. In the close view, the camera is placed in a frontal configuration at 40 centimeters of the localized leaf.

VII. BASIC VERIFICATION OF THE METHOD

The presented robotic leaf-probing strategy assumes that for successful sampling of plant leaves it is advantageous to move first to a closer and fronto-parallel viewing position with respect to the leaf surface. To support this claim, we verify that (i) surface normals of leaves can indeed be accurately

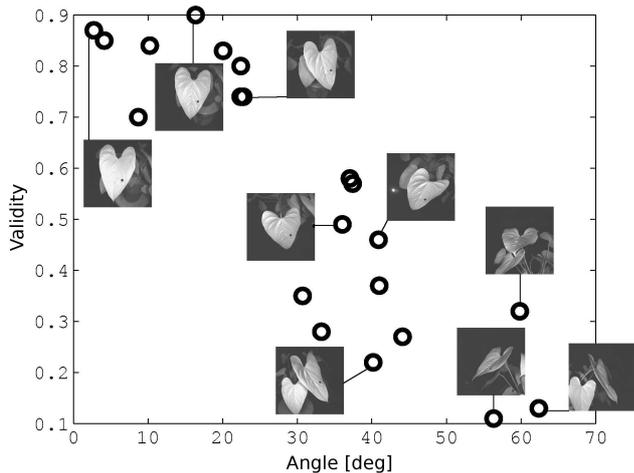


Fig. 5. A. Validity of the leaf as a function of the measured enclosing angle of the surface normal with the z -axis (camera viewing direction) for the artificial leaf. B. The same for a real leaf. The validity measures the correlation between the measured and transformed 2D contour of a segment and a 2D model-leaf contour.

estimated with the given method, and (ii) that moving to a closer, fronto-parallel view of a leaf allows better verification of suitable leaves for probing and thus also a better determination of grasping points.

To test assumption (i), we used a planar artificial leaf. For this purpose, the shape of a real leaf (*Dieffembachia*) was taken and cut from a carton. The shape was also used as a model leaf for this particular experiment. The artificial leaf was attached to a beam and rotated around its center to attain different angles of its surface normal with the viewing direction of the PMD camera. The leaf was rotated in steps of 5 deg, starting at 0 deg, and a depth and an infrared-intensity image was acquired at each step. Using our method, the leaf was segmented and the surface normal was computed by fitting a plane to the 3D points of the segment. The enclosing angle of the measured surface normal with the z -axis in the camera coordinate system was calculated. Fitting of a line to the data revealed an approximate measurement error of about ± 0.7 deg, which demonstrates that a sufficiently accurate estimation of the surface normal can be obtained with the system.

To verify assumption (ii), we used the model-leaf contour to calculate the validity of the extracted segments during the previous experiment. The validity measures the correlation between the measured 2D contour of a segment and a 2D model-leaf contour (see Section V). We observed that the validity decreases with increasing angle, i.e., the further we move away from the fronto-parallel position, the more difficult it becomes to recognize the leaf due to view-dependent shape distortions and other visibility impairments. This also implies that the grasping point cannot be accurately determined past some angle, because the model-leaf contour together with the associated grasping point will fit the segment boundary only very poorly. We further acquired depth and infrared-intensity images from various viewing angles of a real leaf (the very leaf that had also been used to extract the model-leaf

shape in the previous experiment). The computed validities are displayed in Fig. 5 and show the same trend as the validities obtained for the artificial leaf. The experiments demonstrate that the method is capable of extracting the target leaf despite different viewing conditions and without having to change the parameters. A close view for probing is desirable since it increases the amount of data that can be gathered about a leaf (the resolution), which is immediately evident and does not need to be demonstrated.

Combining the data for the artificial and the real leaf showed that the validity follows roughly a linear relationship. By finding the mean distance of the measured points to the fitted line, we obtained an approximate error measure of the computed validity values of about ± 0.1 . However, this measure has been computed for leaves that are mostly planar. Bended or curled leaves might show larger errors.

VIII. LEAF PROBING

After having verified the basic assumptions of our approach, we test the method on different plants and for different viewpoints. Each experiment proceeds in the same way: First, the plant is examined from a far (general) viewing position of the robot arm. The ToF data is processed and a target leaf is selected. Second, using the target's pose, a new robot position is planned and the robot is moved to get a close view of the target. Third, the ToF data from the new view is processed and the target is confirmed if it is of sufficient validity and graspability ($g > -10$). Only then the reaching movement is computed and the grasp is executed. Throughout all the probing experiments, the same set of parameters is used in the algorithms with $d_{\text{merge}} = 1 \text{ cm}^2$. Note that the maximum validities obtained in these experiments are smaller than in the benchmark experiments, since here the model-leaf contour might not perfectly describe the selected leaf due to natural variations in the shape appearance of the leaves.

In Figure 6, the results of the analysis of the ToF data for five experiments (A-E) obtained for the far viewing position are shown. In general, depth segmentation delivered sufficiently good results for identifying targets of interest. Except for experiment C (Fig. 6C), for which segmentation failed, targets of sufficient validity could always be found. The values of the validity and graspability measure are summarized in Table I. Targets selected during the experiments are labeled with a unique number in the figures and the table. The computed grasping points are indicated with a star-shaped symbol in the figures.

Based on the selected target, a close view on the target can be planned using its 3D pose. After moving to the close position, the newly acquired data is analyzed. As can be seen in Figure 7, segmentation improves in the close view compared to the general view, and in all cases except experiment C, target leaves can be confirmed indicated by a sufficiently large validity ($v > 0.3$). However, in experiment E, the validity decreased by 0.14, which is nevertheless still in the error margin of the validity computation. The validity measure can be impaired by many factors, e.g., shape differences of the real leaf compared to the model leaf, non-optimal

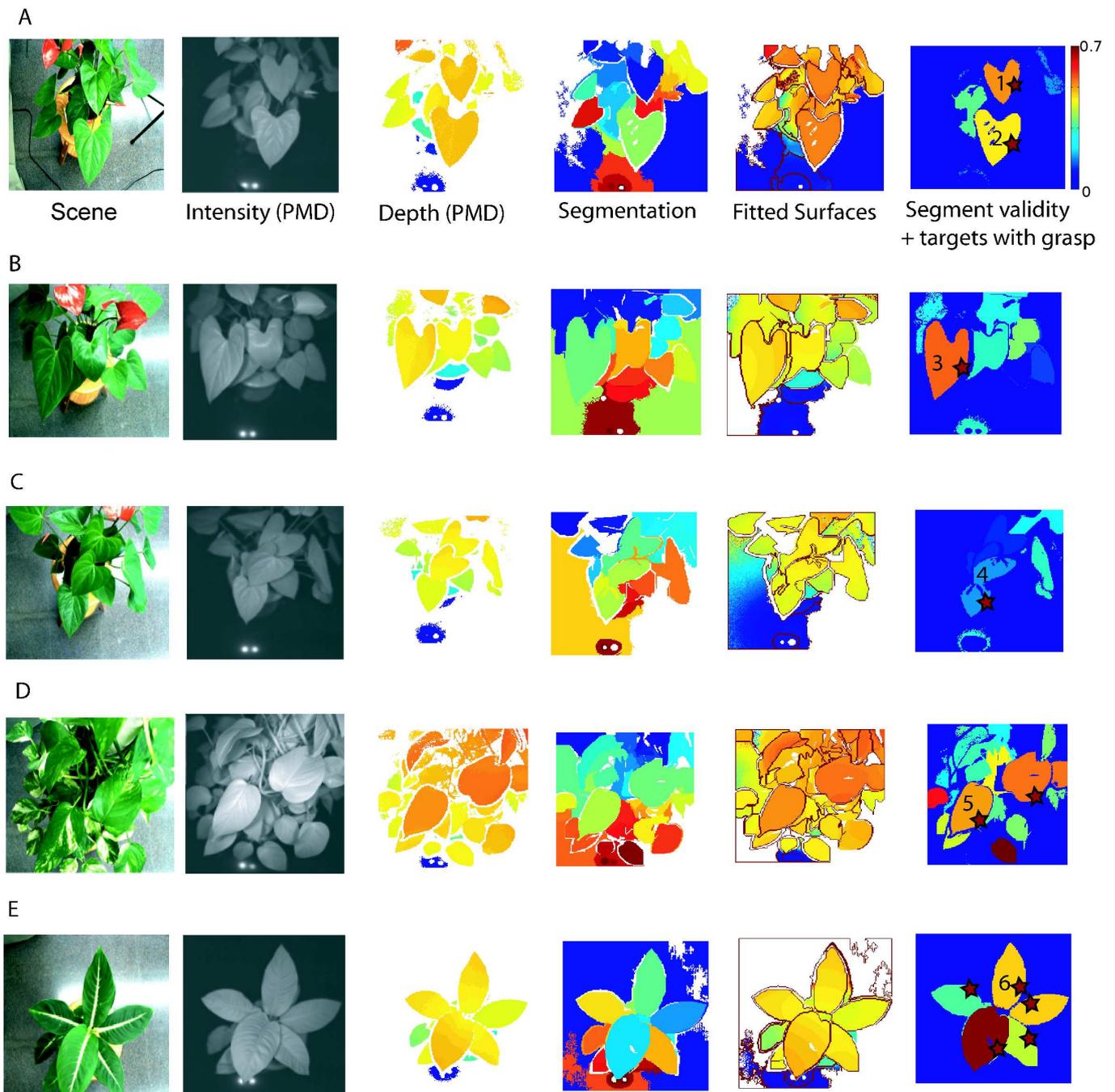


Fig. 6. Segmentation and target selection results for ToF data acquired from a far viewing position relative to the plant for five different experiments (A-E). Depth is color coded with increasing depth coded from red to blue (3rd and 5th column). The grasping points of selected and labeled target segments are marked with a red star. The validity measures the correlation between the measured and transformed 2D contour of a segment and a 2D model-leaf contour.

solutions encountered by the fitting procedure, segmentation errors distorting the boundaries.

Since the contour models are just a rough approximation and the validity estimation is afflicted with some error (see Section VII), validities are expected to increase from the far view to the close view, only if the change in the viewing angle is large (> 30 deg). Graspabilities close to zero indicate that a grasp is executable, however, some noise in the segmentation or data can cause the graspability value to deviate slightly from zero, e.g., experiment D.

The close view in experiment C provides more information on the plant than the far view. Now several leaves can be distinguished, reflected in the larger validity of the segments. As a consequence, a better target (labeled 7) with higher validity can be selected for the grasp.

Furthermore, moving from a far-view to a close-view allows collecting more data about a segment and bringing the camera in a fronto-parallel position with respect to the leaf. In Table I, the enclosing angle of the surface normals with the camera viewing angles of the leaf for the far and the close view are

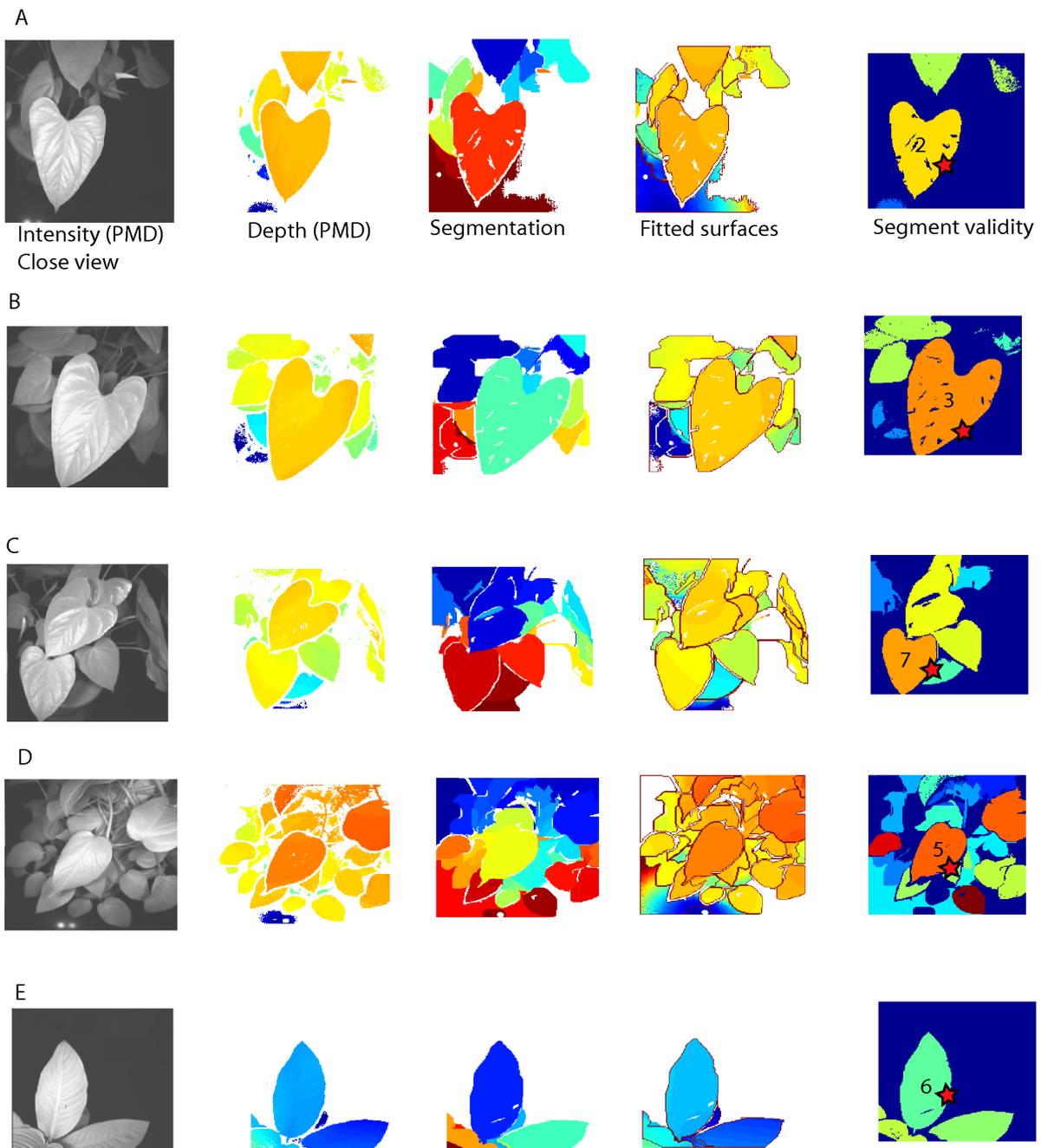


Fig. 7. Segmentation and target selection results for ToF data acquired from a close viewing position relative to the plant for the five different experiments (A-E). Depth is color coded with increasing depth coded from red to blue (2rd and 4th column). The grasp points of selected and labeled target segments are marked with a red star. The results for the close view of target 1 are shown in Fig. 3.

reported, showing that surface normals have been re-estimated after going to the new view and are sufficiently aligned with the z -axis for the close view. In experiment C however the viewing direction of the camera could not be aligned well with the surface normal of the leaf in the close view, because two leaves were merged in the far view and thus no surface normal of a singular leaf could be isolated at the beginning.

The target segments and their respective grasping points are now used to compute the approach for probing the leaf. In Figure 8, the 3D point cloud of target 2 is presented together

with the grasp point (hexagram), the center point of the target (circle), the probing point (diamond), the intermediate goal position (square), the surface normal (black line), and the approach vector connecting all these points (green line). Using this information the intermediate goal position of the robot can be calculated.

The grasps were then executed by first going to the intermediate goal position and then advancing to the probing position. Images of the probing for the experiments are presented in Figure 9A-E, demonstrating the successful execution of the

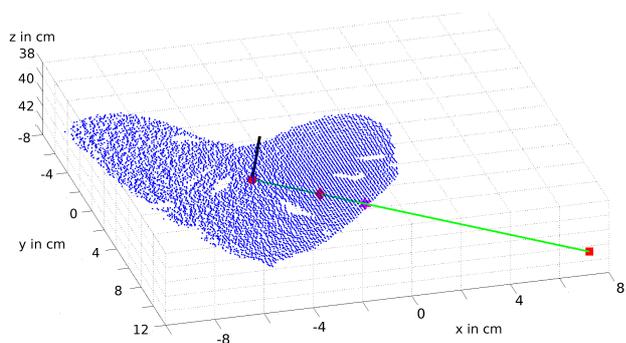


Fig. 8. Target point cloud and approach points. The 3D points of the selected target segment 2 of experiment A (see Figs. 6-7A) are shown together with the associated center point (circle), probing point (diamond), grasping point (hexagram), the intermediate goal position (square), the surface normal (black line), and the approach vector (green line), which intersects all these points. Distances are given in centimeters.

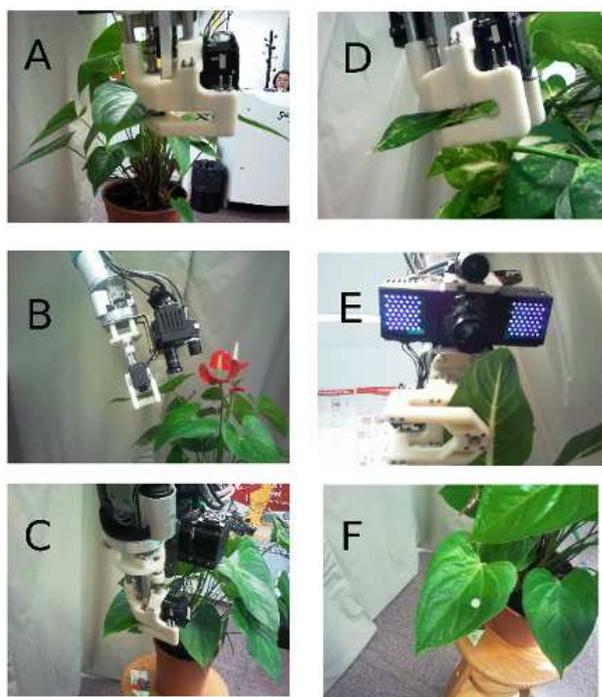


Fig. 9. A-E. Color images documenting the successful execution of grasps. The probing tool could be accurately placed on the leaf and a disc-shaped piece of the leaf could be cut. F. A leaf after a sample has been taken with the cutting tool.

task. Once the cutting tool is correctly placed, it can be used to take a small sample of the leaf by cutting out a small disc. In Figure 9F an image of a leaf after sampling is shown.

The accurate placement of the probing tool indicates that surfaces have been correctly estimated by the approach. The successful execution can be partly attributed to a precise leaf estimation, i.e., surface normal and grasping points, which could be obtained using the data acquired from the close view position.

Movies of the experiments can be found at <http://www.iri.upc.edu/people/galenya/pub/LeafProbing>.

TABLE I

VALIDITY, GRASPABILITY, AND ENCLOSED ANGLE α IN DEG OF THE SURFACE NORMAL WITH THE CAMERA VIEWING DIRECTION FOR THE FAR VIEW AND THE CLOSE VIEW OF EXPERIMENTS A-E IN COMPARISON.

Example	$v_{\text{view 1}}$	$v_{\text{view 2}}$	$g_{\text{view 1}}$	$g_{\text{view 2}}$	$\alpha_{\text{view 1}}$	$\alpha_{\text{view 2}}$
A (target 1)	0.51	0.48	-31	0	6.9	4.7
A (target 2)	0.44	0.46	0	0	5.7	6.5
B (target 3)	0.55	0.51	0	0	18.5	7.3
C (target 4/7)	0.16	0.5	0	0	15.6	17.0
D (target 5)	0.5	0.54	0	-5	1.2	2.6
E (target 6)	0.47	0.33	-29	0	26.8	6.0

IX. CONCLUSIONS

We presented a method for modeling, monitoring, and sampling plant leaves using infrared-intensity images and depth maps acquired with a PMD camera. Since quadratic surface models are used to guide the segmentation of the infrared-intensity image, sparse or noisy depth data can be used, which often poses a problem to approaches working in the depth space directly. Then, segments that are candidates to being leaves are ranked and a closer view of the most promising is taken. In this way, for example, two leaves that were initially merged into a single segment due to insufficient resolution in the far view, could be separated and modeled from the close view. Grasping points could be extracted with high accuracy and disc samples of leaves were successfully cut.

The problem of leaf segmentation has been addressed before by Quan *et al.* (2006), who proposed an image-based plant modeling system based on structure from motion, but which requires user interaction in the segmentation procedure to delineate some leaves [10]. In another related work, leaves were segmented from combined color images and stereo depth, and subsequently classified using the normalized centroid contour distance [9]. Different from these approaches, we extract leaves from ToF data and infrared-intensity images. Segmentation is fully automatic and based on a novel depth-segmentation algorithm which can be applied to sparse or noisy depth data and cope with curved surfaces. Another difference is that leaf models are fitted explicitly which allows localizing grasping points.

The proposed system for automated plant probing is related to vision-based robotic systems for fruit and vegetable picking, which have been proposed in the past for the automation of harvesting tasks [6]. Commonly these systems first process and segment the data in order to identify and represent the target. Based on this representation, a robot action, i.e., cutting or grasping, is executed. Often the image processing task is eased by fixing the environment in a specific manner. For example, in a fruit-detachment system developed by Feng *et al.* (2008) strawberries were grown on a uniformly colored surface to simplify image segmentation [7]. In our system the environment is less constrained, and the proposed computer-vision system is thus more complex. Furthermore, a new robotic application, i.e., the automatic sampling of leaves with a specific cutting tool, was introduced and explored. To the authors' knowledge, this is the first time that an active vision

approach using ToF depth has been applied to robotized plant measuring.

The method is based on several assumptions: (i) the boundaries of leaves are visible in the infrared-intensity image, (ii) the leaf surfaces can be modelled by a basic quadratic function, (iii) leaves of a specific plant type can be described by a common 2D contour, (iv) leaves are large enough to allow analyzing them with a ToF camera, and (v) the leaves are static during probing. These assumptions may be violated under certain conditions, but nevertheless we expect the method to be applicable to many different types of plants given a controlled environment.

In conclusion, we tackled a quite complex task that required the extraction of task-relevant plant parameters from plant images using a multi-stage algorithm, as well as the difficult problem of the actual execution of the robot motion towards the plant. The automation of plant probing has potentially a wide range of applications both in the agricultural industry where certain, currently manual, tasks have to be executed repetitively for many plants, usually of the same type, and in botanic experimentation, e.g., for phenotyping, where leaf sample discs are widely used to analyze plant development in order to determine the factors affecting growth.

ACKNOWLEDGEMENTS

This research is partially funded by the EU GARNICS project FP7-247947, the project PAU+ (DPI2011-27510), and the Grup consolidat SGR155. B.Dellen acknowledges support from the Spanish Ministry of Science and Innovation through a Ramon y Cajal program.

REFERENCES

- [1] A. Kolb, E. Barth, and R. Koch: "ToF-sensors: New dimensions for realism and interactivity", IEEE CVPR Workshops, vol. 1-3, Anchorage, pp. 1518-23, June 2008.
- [2] A. Nealen, M. Muller, R. Keiser, E. Boxerman and M. Carlson: "Physically Based Deformable Models in Computer Graphics", Computer Graphics Forum, Vol. 25(4), pp. 809-6, 2006.
- [3] R.B. Rusu, A. Holzbach, R. Diankov, G. Bradski and M. Beetz: "Perception for Mobile Manipulation and Grasping using Active Stereo", 9th IEEE-RAS Intl. Conf. on Humanoid Robots, pp. 632-638, Dec. 2009.
- [4] T. Fourcaud, X. Zhang, A. Stokes, H. Lambers and Ch. Köner: "Plant Growth Modelling and Applications: The Increasing Importance of Plant Architecture in Growth Models", Annals of Botany 101: 1053-1063, 2008.
- [5] E.J. Van Henten, B.A.J. Van Tuijl, G.-J. Hoogakker, M.J. Van Der Weerd, J. Hemming, J.G. Kornet, and J. Bontsema, "An Autonomous Robot for De-leafing Cucumber Plants grown in a High-wire Cultivation System", *Biosystems Engineering* 94(3), pp. 317-323.
- [6] T. Grift, Q. Zhang, N. Kondo, K.C. Ting, "A review of automation and robotics for the bio-industry". *Journal of Biomechanics Engineering*, vol. 1, pp. 37-54, 2008
- [7] G. Feng, C. Qizin, and M. Masateru, "Fruit detachment and classification method for strawberry harvesting robot", *International Journal of Advanced Robotic Systems*, vol 5, pp. 41-48, 2008
- [8] S. Kitamura and K. Oka, "Recognition and cutting system of sweet pepper for picking robot in greenhouse horticulture", *Mechatronics and Automation, 2005 IEEE International Conference*, vol. 4, pp. 1807-1812, 2005
- [9] Ch.-H. Teng, Y.-T. Kuo, and Y.-S. Chen, "Leaf segmentation, classification, and three-dimensional recovery from a few images with close viewpoints". *Optical Engineering* 50(3), doi:10.1117/1.3549927, 2011.
- [10] L. Quan, P. Tan, G. Zeng, L. Yuan, J. Wang and S.B. Kang: "Image-based Plant Modelling", ACM Siggraph, pp. 599-604, 2006.
- [11] Y. Song, R. Wilson, R. Edmondson and N. Parsons: "Surface Modelling of Plants from Stereo Images", 6th IEEE Intl. Conf. on 3D Digital Imaging and Modelling (3DIM 2007).
- [12] G. Taylor and L. Kleeman: "Robust Range Data Segmentation Using Geometric Primitives for Robotic Applications", 5th Iasted Int. Conf. on Signal and Image Processing 2003.
- [13] B.I. Loch, J.A. Belward, and J.S.Hanan, "Application of Surface Fitting Techniques for the Representation of Leaf Surfaces". In Zerger, A. and Argent, R.M. (eds): *MODSIM 2005 International Congress on Modelling and Simulation*, pp. 1272-1278, December 2005.
- [14] R. Klose, J. Penlington and A. Ruckelshausen: "Usability study of 3D Time-of-Flight cameras for automatic plant phenotyping", Workshop on Computer Image Analysis in Agriculture, pp. 93-105, August 2009.
- [15] S. Foix, G. Alenyà, J. Andrade-Cetto and C. Torras: "Object modeling using a ToF camera under an uncertainty reduction approach". IEEE Intl. Conf. on Robotics and Automation (ICRA'10), Anchorage, pp. 1306-1312, 2010.
- [16] B. Bartczak and R. Koch: "Dense Depth Maps from Low Resolution Time-of-Flight Depth and High Resolution Color Views", Lecture Notes in Computer Science 5876, pp. 228-239, 2009.
- [17] A. Bleiweiss and M. Werman: "Fusing time-of-flight depth and color for real-time segmentation and tracking", Workshop on Dynamic 3D Imaging (Dyn3D'09), pp. 58-69, 2009.
- [18] F.F. Khalil and P. Payeur: "Dexterous Robotic Manipulation of Deformable Objects with Multi-Sensory Feedback - a Review", in "Robot Manipulators Trends and Development", edited by A. Jimenez and B.M. Al Hadithi, INTECH, pp. 587-619, March 2010.
- [19] S. Foix, G. Alenyà, and C. Torras: "Towards plant monitoring through Next Best View". *Proc. Intl. Conf. of the Catalan AI Society (CCIA'11)*, Lleida, 2011.
- [20] G. Alenyà, B. Dellen, and C. Torras: "3D modelling of leaves from color and ToF data for robotized plant measuring". *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA'11)*, Shanghai, May 2011.
- [21] B. Biskup, H. Scharr, A. Fischbach, A. Wiese-Klinkenberg, U. Schurr and A. Walter, "Diel Growth Cycle of Isolated Leaf Discs Analyzed with a Novel, High-Throughput Three-Dimensional Imaging Method Is Identical to That of Intact Leaves". *Plant Physiology* 149, pp. 1452-1461, March 2009.
- [22] S. Foix, G. Alenyà, and C. Torras, "Lock-in Time-of-Flight (ToF) cameras: a survey," *IEEE Sensors J.*, vol. 11(9), pp. 1917-1926, 2011.
- [23] S. Fuchs and G. Hirzinger, "Extrinsic and depth calibration of ToF-cameras," in *Proc. 22nd IEEE Conf. Comput. Vision Pattern Recog.*, vol. 1-12, Anchorage, June 2008, pp. 3777-3782.
- [24] S. Fuchs and S. May, "Calibration and registration for precise surface reconstruction with time of flight cameras," *Int. J. Int. Syst. Tech. App.*, vol. 5, no. 3-4, pp. 274-284, 2008.
- [25] S. Foix, G. Alenyà, and C. Torras, "Towards plant monitoring through next best view," in *Proc. 14th Int. Conf. Catalan Assoc. Artificial Intell.*, Lleida, Oct. 2011.
- [26] P.F. Felzenszwalb and D.P. Huttenlocher: "Efficient graph-based image segmentation". *International Journal of Computer Vision* 59, pp. 2004, 2004.
- [27] J.B. Kruskal, "On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem", *Proceedings of the American Mathematical Society*, vol. 7, no. 1, pp. 4850, 1956