3D Sensor planning framework for leaf probing

Sergi Foix, Guillem Alenyà and Carme Torras

Abstract— Modern plant phenotyping requires active sensing technologies and particular exploration strategies. This article proposes a new method for actively exploring a 3D region of space with the aim of localizing special areas of interest for manipulation tasks over plants. In our method, exploration is guided by a multi-layer occupancy grid map. This map, together with a multiple-view estimator and a maximum-information-gain gathering approach, incrementally provides a better understanding of the scene until a task termination criterion is reached.

This approach is designed to be applicable for any task entailing 3D object exploration where some previous knowledge of its general shape is available. Its suitability is demonstrated here for an eye-in-hand arm configuration in a leaf probing application.

I. INTRODUCTION

Plant phenotyping studies the influence of environmental factors on the observable traits of plants. The success of such studies depends on the data extracted from a series of long-term monitoring experiments over a great number of plants under multiple environmental conditions. Measures can be obtained in two different scenarios. The first one includes regular fields and mobile sensors, either mounted on aerial vehicles using remote sensing techniques [1], or on ground robots [2]. Obviously, climate conditions can not be controlled here. The second one uses greenhouses, where variables like temperature, humidity, and light, can be controlled. The common setup includes large greenhouses with several isolated zones, and conveyor belts that carry each plant from its sitting position to a measure chamber where a rich set of sensors takes measurements before returning them [3]. The throughput obtained in such installations is considerably high. However, sensors in the measuring chamber are in a pre-defined position, and measurements are sometimes obtained from an inadequate point of view. Additionally, one of the main limitations is the difficulty to measure or perform actions that require contact with the plant, such as chlorophyll measurement or the extraction of disk samples for DNA analysis [4].

Therefore, a major step forward is to provide the system with the ability to move its perceptual unit, so that it can naturally adapt to the characteristics of each plant. In this paper a robotic system is proposed to overcome this weakness, that involves a time-of-flight camera (ToF) and a probing tool mounted on the end-effector of a robot manipulator. The approach includes three steps: (i) selection of the target leaf from a plant, (ii) exploration of this leaf to gather enough

The authors are with the Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain, {sfoix, galenya, torras}@iri.upc.edu.
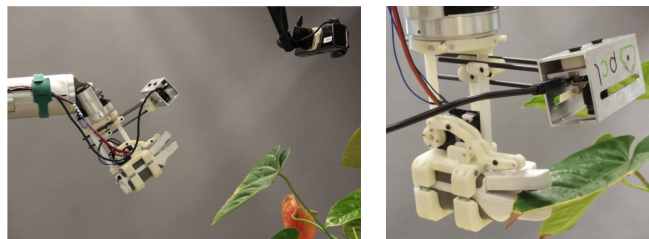
(a) Experimental setup   (b) Leaf probing action

Fig. 1: (a) Overall view of the complete setup, with the robot carrying the ToF camera and the tool, and the Kinect camera. (b) Observe that the leaf probing task requires the clearance (above and below) of a sector of the leaf.

information, and (iii) effective execution of probing. The first step is usually specified by a botanical expert that defines a criterion to choose the leaf, for example the biggest one, or the i-th leaf from the stem. The last step, the measuring action, has been already presented in previous works [4]. This paper focuses on a method for solving the second step.

We propose to perform task-driven exploration, that is gathering the necessary information to execute a task, with a method where the task is encoded in a combination of 3D occupancy grid maps. Although we concentrate on spatial restrictions like clearance, it is also shown that other task constraints, like veins and yellow spots that have to be avoided, can be represented in additional occupancy grid maps following the same idea. The goal of task-driven exploration is to move the robot through a sequence of views that will contribute to maximising the information for solving the task. We propose an algorithm that uses an information-gain approach to compute the expected benefit of each new possible view, and combines it with other aspects, such as the proximity to the current view, to obtain the best next view at each iteration. Observe that this is a local approach, and that it cannot be optimal since the next position of the sensor only depends on the available information at each iteration. An optimal solution would require a complete model of the plant beforehand. That is not feasible in plant phenotyping, since plants from the same species are quite different and even the same plant changes largely over time.

The rest of the paper is organized as follows. Related works are discussed in Section II. In Section III, our leaf probing exploration framework is explained. Experimental results are presented and discussed in Section IV. Finally Section V is devoted to conclusions.
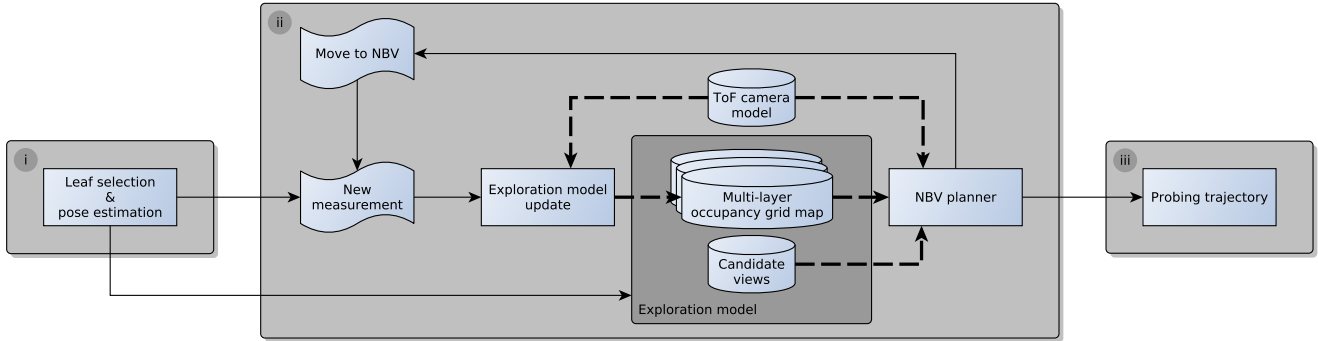
Fig. 2: 3D sensor planning framework.

II. RELATED WORK

The ability to move the sensor and consequently provide a suitable view offers several advantages and it has been investigated both in the field of mobile robotics and in the field of active sensors mounted onto robot manipulators [5]. Regarding manipulators, active sensor planning algorithms can be divided into two main groups depending on the available initial information: model and non-model based.

Model-based methods require to have a previous complete model of the object, usually a CAD model, and are useful for industrial inspection or part localization. These approaches are often based on visual servoing techniques [6]. Here, planning algorithms find optimal solutions and consequently, execute the minimum number of views to complete the task.

Non-model based methods assume vague knowledge about the object. Using partial information the algorithm has to produce a next-best-view (NBV) to incrementally gather new data. This approach is commonly used for object modeling [7] and object recognition [8]. Necessarily, since NBV is determined based on partial knowledge, an optimal solution cannot be guaranteed. Our approach fits into this group.

Classical non-model based approaches rely on frontier-based exploration. The camera is iteratively moved relative to the closest or largest visible frontier, or among the two, to the one with the easier configuration for the robot [9]. Their implicit objective is the complete exploration of an area, it being difficult to explicitly enforce a different task. These methods are very adequate when using accurate sensing and positioning devices [10], but not so much in the presence of noise, since they do not handle uncertainty naturally. In special cases though, where revisiting actions can be ensured, the accumulated noise in the scene can be improved through uncertainty reduction approaches [11].

Information gain is one of the most used measures for candidate view selection. Each possible view, when visited, provides a certain amount of new information. Even revisiting already captured areas can provide newer information, specially if observed from a different point of view. To estimate the gain for each candidate view, some predictions about the unexplored regions must be made. A typical approach is to make assumptions about the nature of the already captured surfaces. For instance, adjusting B-Splines

to sections of the 3D object [12], or modeling the surface as a combination of a trend and some disturbances [13].

Note that, while a coarse model of a plant's leaf can be feasibly obtained for our application, it can not be easily exploited by the current approaches. Instead, it has been shown that 3D occupancy grid maps are well suited for incremental object modeling [14], [15]. Therefore, we propose a method that takes advantage of both a coarse model of a leaf and 3D occupancy grid maps to graphically encode the task of finding a good probing point.

III. LEAF PROBING FRAMEWORK

Neither leaf probing nor chlorophyll measurement requires to have a complete leaf's model to accomplish its goal. Instead, only specific regions in the leaf's contour are needed. To specify the task, we consider two types of information: the prior knowledge of the task and the on-line data, both codified using probabilistic occupancy maps. This section emphasises the following three main ideas:

1) All required data can be represented within a multi-layer occupancy map, where each layer codifies relevant information that is semantically different (Sec. III-C.1). Particularly, the space occupied by the leaf and the free space for allowing the tool to reach the leaf.
2) The Information Gain ($\mathcal{IG}$), used as a criterion to evaluate potential new views, can be easily defined and computed from the multi-layer representation (Sec III-D). Its correct computation requires an accurate model of the ToF camera (Sec. III-B).
3) The task termination criterion signals when enough data is available to perform the task. Our representation facilitates its definition and evaluation (Sec. III-C.1).

To easily understand the entire framework, we will first describe the global set-up and thereafter its modules and how they are interconnected. See Fig. 2 for an overall view of the system. We stress again that our method focuses on the main step of exploration planning between consecutive views until task termination (ii), and not on the initial step of leaf selection and pose estimation (i), nor on the final probing trajectory approach (iii).

(a) Initial state of the system.

(b) Exploration model.

(c) Zenithal measurement update.

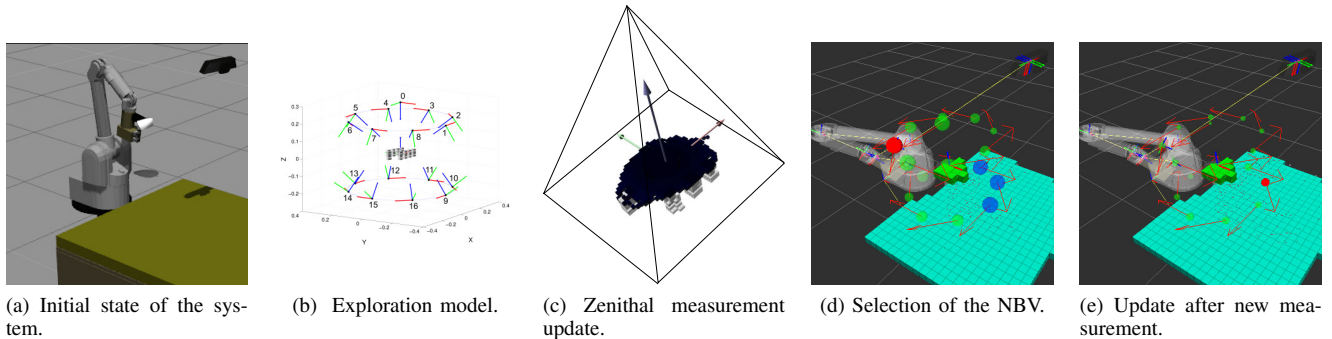(d) Selection of the NBV.

(e) Update after new measurement.

Fig. 3: Step-by-step graphical interpretation of the overall procedure. (a) The system assumes having a leaf in the field-of-view of the camera. Once the leaf is detected, the exploration model is introduced into the system (b), together with the first measurement (c). Views are iteratively selected depending on their expected $\mathcal{IG}$, proximity and reachability (d). Green spheres represent candidate viewpoints that have not been selected yet as a Next-Best-View, blue spheres indicate already selected but not reachable viewpoints and the red sphere shows the one that has been selected and is being evaluated. The radius of the sphere represents the expected $\mathcal{IG}$. Observe how, after a new measure is integrated, candidate expected $\mathcal{IG}$s get reduced (e).

A. Set-up

The experimental set-up consists of a Barrett WAM arm (robot manipulator) and three sensors: a PMD Camboard (ToF camera), a spad meter (chlorophyl measuring tool), and a Kinect camera (see Fig. 1a). The ToF camera is, in conjunction with the spad meter, rigidly attached to the robot's end-effector, in such a way that permits controlling the robot for both capturing detailed views from informative regions of interest, and taking chlorophyll measurements from selected target leaves (see Fig. 1b). The RGB-D camera is deliberately situated on the ceiling, in a zenithal configuration, to allow a complete overall view of the scene. Its main purpose is to feed the obstacle avoidance map so that safe trajectories can be successfully planned.

B. ToF camera model

It is important to highlight the relevance of having a good characterization of the depth sensor's model. The sensor's model is not just used for updating the occupancy grid maps but also for computing the information gain estimation when simulating the candidate views (Sec. III-D). Note that in ToF cameras, the uncertainty associated to each pixel is different.

Due to their technology, ToF camera's depth measurements have attached a set of associated errors (see [16] for a comprehensive list). After calibration and filtering, the remaining errors can be approximated by Gaussian noise with zero mean and uniform standard deviation and independent of any other measurement. The calibration process is long, tedious, and has to be performed for multiple distances. The common approach we follow is to define a 10 cm. safety range distance and perform the calibration within that range. For the experiments in this paper we have selected a preferred distance of 35 ± 5 cm.

C. Exploration model

Once a leaf has been selected and its pose correctly estimated (Fig. 3a), an initial task-driven *exploration model*

is defined (Fig. 3b). Its aim is to incrementally encode knowledge about the scene to effectively fulfill the task (i.e. leaf probing). The exploration model is composed of a *multi-layer occupancy grid map* and a *set of candidate viewpoints*, see Fig. 4 for a schematic representation.

1) Multi-layer occupancy grid map: In a similar way to Lu *et al.* [17], we have subdivided the occupancy representation of the exploration model into three semantically-separated layers $\{m^{\mathbf{task}}, m^{\mathbf{state}}, m^{\mathbf{obs}}\}$. By doing this, our method obtains a wide versatility that facilitates four key aspects: the specification of the task halting criterion, the precise adjustment of particular exploratory attractors, the correct treatment of possibly occluded regions of interest and the computation of obstacle avoidance trajectories.

The aim of the *task termination* layer $m^{\mathbf{task}}$ is twofold: first, to indicate whether the exploration can already be halted; and second, to act as a prior for the NBV planner (Section III-D). This layer is composed of what we call regions of interest (ROIs), e.g. in Fig. 5. Each ROI is defined as a region of expected occupancy in the model, and acts, by itself, as a global halting criterion. That means that if the expected occupancies within a certain ROI are fulfilled after a measurement, the exploration task has finished and a probing trajectory can be carried out. Our leaf probing model is composed of 9 separated ROIs wisely located at the edges of the estimated shape of the leaf (Fig. 5a). Each ROI is composed of three bounding boxes or here called *bricks*, two of them labelled as free, one at the top and one at the bottom, and another one labelled as occupied in the middle (Fig. 5b). It is very intuitive to see how these ROIs represent the desired open and occupied spaces at the leaf edges that can allow the probing tool to take a measurement. Notice that we do not need to characterize the complete occupancy model of a leaf but only those parts (ROIs) that help to solve the probing task.

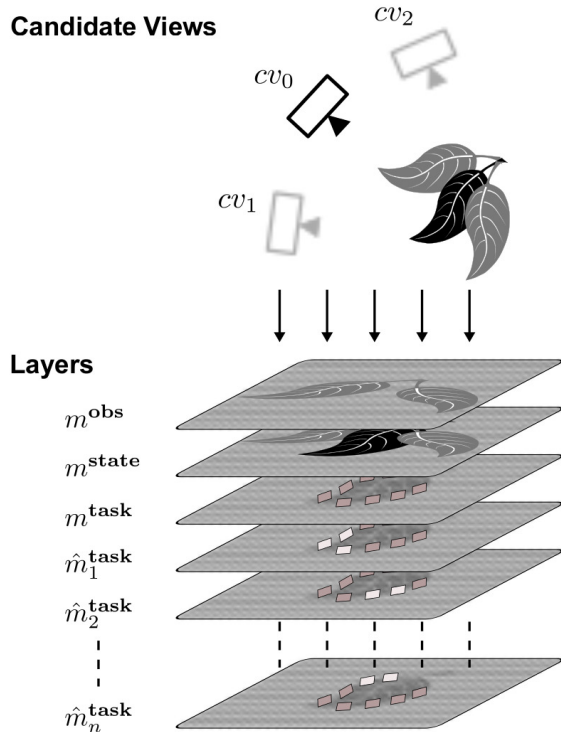The *state* layer $m^{\mathbf{state}}$ is the one keeping the complete

Fig. 4: Schematic view of the multi-layer occupancy grid map. Each layer is a 3D occupancy map that semantically separates different informations. The potential information-gain obtained from the different views cv_i is estimated from this multi-layer grid map.

update of all measurements taken during an experiment. As a result, and in conjunction with the $m^{\mathbf{task}}$ layer, the NBV planner can thereafter predict a more realistic estimation of the information gain. Such prediction is accomplished by simulation, i.e., every candidate viewpoint is ray-traced over $m^{\mathbf{state}}$. Once simulated, each new virtual measurement is updated into a copy $\hat{m}_i^{\mathbf{task}}$ of the global $m^{\mathbf{task}}$ layer for posterior computation of its expected $\mathcal{IG}$ (Section III-D).

The *obstacle avoidance* layer $m^{\mathbf{obs}}$, on the other hand, is dedicated to represent the clearance working space of the robot. This allows the NBV planner to return safe and collision-free trajectories. In our approach, this layer is initialized at the beginning of each experiment with an overall view of the scene coming from a RGB-D sensor located in a vantage point at the ceiling; it can also be updated with every new measurement acquired from the eye-in-hand depth sensor, if higher precision is requested.

2) Candidate viewpoints: A pre-defined set of vantage points $\mathcal{C} = \{\, cv_i \,|\, i = 1, ..., k \,\}$ must be chosen in order to guarantee a good coverage for all ROIs within the $m^{\mathbf{task}}$ layer, see Fig. 3b. Albeit locally pre-defined offline, both the *multi-layer occupancy grid map* and the set of *candidate viewpoints* are updated online. At this moment we assume confidence on the robot's pose, and therefore *candidate*
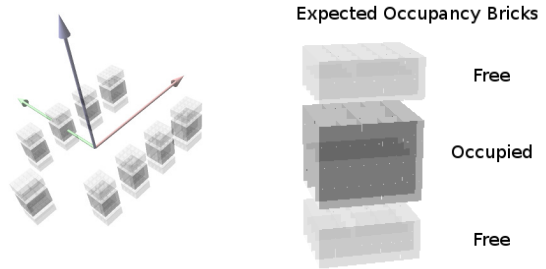
(a) Task termination layer. (b) Bricks within one ROI.

Fig. 5: (a). Graphical representation of the 9 ROIs within the task termination layer. (b) Composition of *bricks* into a single ROI.

viewpoints are updated just once, based on the initial leaf pose estimation. In the future, with the aim of increasing accuracy, it is planned to add the uncertainty of the robot within the system and to consequently incorporate the leaf's pose estimation module into the main loop and adapt the views accordingly.

D. Next-Best-View planner

Prior to the selection of the NBV and after the last measurement update, the system checks within the $m^{\mathbf{task}}$ layer for triggered probing termination ROIs. If any of them is active, meaning that their expected free-occupied-free preconditions are fulfilled, the probing trajectory to its corresponding grasping point will be executed. On the contrary, if none of the ROIs satisfies the halting criterion, the planner will compute the next-best-view cv^* based on the $\mathcal{IG}$ cost function.

Our approach is based on the multi-objective performance criterion described by Mihaylova *et al.* in [18] and has the following form:

$$cv^* = \operatorname*{arg\,min}_{cv_i \,\in\, J} \|cv_i\| \;, \tag{1}$$

where $\| \; \|$ is the euclidean distance from the current view and J is the set containing the 10% of views with the highest expected information gains:
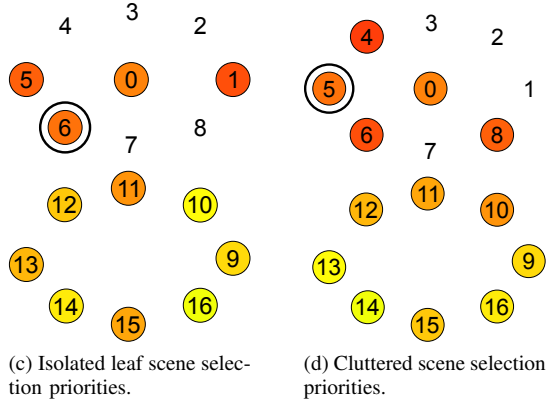
$$\mathcal{IG} = \sum_o \Delta H(\hat{m}_{i,o}^{\mathbf{task}}) + \beta \sum_f \Delta H(\hat{m}_{i,f}^{\mathbf{task}}) \;, \tag{2}$$

where $\Delta H(\hat{m}_{i,o}^{\mathbf{task}})$ and $\Delta H(\hat{m}_{i,f}^{\mathbf{task}})$ are, respectively, the *expected information gains* of occupied and free bricks when going from the current $m^{\mathbf{task}}$ to the simulated $\hat{m}_i^{\mathbf{task}}$ layer; and β allows to balance their relative contributions. While β is very relevant in tasks where a type of brick is more critical than the other, given the equal influence of both types in the probing task, this parameter is set to 1. Note that Eq. 2 can be also interpreted as a weighted *change of entropy* between prior and posterior probability density functions.

Once the system has chosen the most informative and reachable candidate view, a new measurement is taken (Fig 3d) and $m^{\mathbf{state}}$ is updated (Fig 3e).
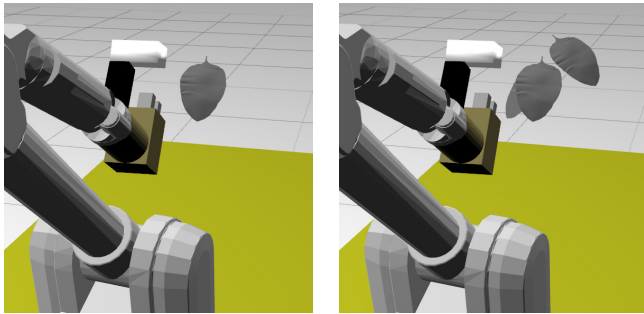
(a) Simulated isolated leaf scene.

(b) Simulated cluttered scene.

(a) Real isolated leaf scene.

(b) Real cluttered scene.

Fig. 7: Single and cluttered scenarios for the real experiment

(c) Isolated leaf scene selection priorities.
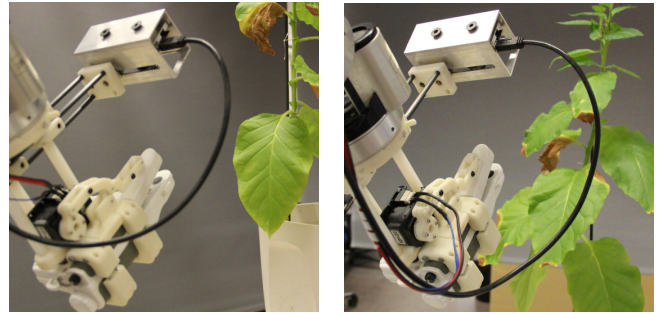

(d) Cluttered scene selection priorities.

Fig. 6: (c) and (d) show the priority of views in scenes (a) and (b) respectively. Each number refer to a candidate view, see Fig. 3b. Colours indicate priority, starting at yellow and finishing at red. The most selected view after all the experiments is marked with an extra-circle. These data have been collected from 100 simulations per scene.

TABLE I: Simulation (100 Exp.)

N. of views	¬obstacle	obstacle
2	64.21%	36.26%
3	28.42%	61.54%
4	5.26%	2.20%
5	1.05%	0%
6	1.05%	0%

Notes: Percentage of experiments that finished in a certain number of views. Some experiments resulted inconclusive due to external errors in the path planner (5 in ¬**obstacle** exp. and 9 in the **obstacle** exp.)

IV. EXPERIMENTS

Figure 3 describes graphically the step-by-step procedure of the proposed planning framework. The experiments have been carried out under simulation and subsequently tested, using the very same algorithm and parametrization, over the real robot in a very similar scenario, see Figs. 6 and 7.

A. Simulation

This experiment analyses the behaviour of the system in an isolated leaf and multiple-leaves scenes, Fig. 6a and 6b respectively. The goal is to first evaluate the suitability of the method on both tidy and cluttered scenes, and then compare the robot behavior between simulation and reality in Sec. IV-B. Figures 6c and 6d display the priorities of views while planning for each of the experiments. For clarity, the candidate viewpoints considered have been re-arranged in a 2D graph. Note that $[0..8]$ correspond to the top views and $[9..16]$ correspond to the bottom views. It can be seen that, in both experiments, the planner tries to go to the most informative views first, those that point behind the leaf. We have purposely defined these points to be non-reachable by the robot. Observe, for instance, the behavior in the cluttered scenario: the planner chooses, in the first place, view 13 as

the closest view among the 10% of those with the highest IG; since it is unreachable, this view is removed from the candidate views list and following the same criterion the NBV is chosen again until a reachable view is obtained. The final solution for both experiments is a top lateral-left view (5 or 6).

Table I shows how the presence of obstacles affects the system. In the isolated leaf scene, most of the experiments (64%) fulfill the task termination conditions at the second view. This is good news for plant phenotyping where high throughput is required. Observe that when obstacles are present, the task is finished with a second view 36% of the times, and requires one more view 61% of the times.

B. Real

We carried out a set of 20 real experiments, using the same algorithm and parametrization than in the simulation, and with very similar scenes, see Fig. 7. Half of the experiments are devoted to the isolated leaf scene, and half to the cluttered scene. In both cases the robot behaved in the same way as the simulated ones, and it tried to go to the more informative views, and when not possible it chooses the NBV until probing was successfully accomplished. The main difference is that in a cluttered situation the view number 6 is preferred 80% of the times instead of number 5, which is the one preferred in simulation. However, both views are close and the difference in the obtained gain is small. [1]

[1] Additional material at:
http://www.iri.upc.edu/groups/perception/leafProbing.

V. Conclusions

A solution for plant phenotyping has been presented, composed of a manipulator robot carrying a ToF camera and a specialized probing tool. Although our experiments can include either a tool for chlorophyll measuring or a probing tool for leaf sampling, both tasks involve the same framework: the robot changes the point of view of the camera to take new images, and when enough information is acquired the task is performed. None of the tasks requires the complete model of the leaf but just to view a small part of the leaf and a clearance zone.

In this article a novel 3D task representation based on a multi-layer map has been introduced, where each layer codifies relevant information that is semantically different. It has been shown that this representation has three main advantages. First, it allows to define tasks based on prior information available using a combination of free and occupied space, and even possible constraints.

Second, a formulation has been introduced to compute the expected information gain from this representation. The correct computation requires the accurate calibration of the ToF camera because of the uncertainty in depth related to each pixel. It has been also shown that $\mathcal{IG}$ can be effectively used to select the next-best-view. As the application may require to minimize the motion of the robot, a criterion has been introduced to prefer closer views even if they provide slightly less information.

Third, this representation allows the natural specification of the task termination conditions, which are the minimum units of information required to enable the execution of the task.

We have purposely left out of the scope of this paper the generation of the list of possible views, and we have considered that is provided. In the future, and following our interest in taking into account the task, we will explore the generation of the list of views to contain only *interesting* views. We envisage that the evaluation of its interest can be done off-line using simulations with the presented framework, and something similar to lookup tables of relative positions depending on the task can be created.

Finally, the practical experiments have revealed that the relative position of the robot and the plant is important, as some of the views are not reachable. A common approach in automated plant phenotyping is to control the orientation of the pot containing the plant. Although this adds a new degree-of-freedom to the control of the robot, it can be easily integrated in the proposed framework.

References

[1] I. Colomina and P. Molina, "Unmanned aerial systems for photogrammetry and remote sensing: A review," {*ISPRS*} *Journal of Photogrammetry and Remote Sensing*, vol. 92, no. 0, pp. 79 – 97, 2014.

[2] S. Nielsen, K. Jensen, A. Bøgild, O. J. Jørgensen, N. Jacobsen, C. L. Jæger-Hansen, and R. N. Jørgensen, "A low cost, modular robotics tool carrier for precision agriculture research," in *11th Int. Conf. on Precision Agriculture*, 2012.

[3] B. Berger, "High-throughput phenotyping of cereals at the plant accelerator," in *Plant and Animal Genome XXI Conf.* Plant and Animal Genome, 2013.

[4] G. Alenyà, B. Dellen, S. Foix, and C. Torras, "Robotized plant probing: Leaf segmentation utilizing time-of-flight data," *Robotics Autom. Mag.*, vol. 20, no. 3, pp. 50–59, 2013.

[5] S. Chen, Y. Li, J. Zhang, and W. Wang, *Active sensor planning for multiview vision tasks*. Berlin Heidelberg: Springer-Verlag, 2008.

[6] F. Kececi, M. Tonko, H. Nagel, and V. Gengenbach, "Improving visually servoed disassembly operations by automatic camera placement," in *Proc. IEEE Int. Conf. Robotics Autom.*, vol. 4, Leuven, May 1998, pp. 2947–2952.

[7] W. R. Scott, G. Roth, and J. F. Rivest, "View planning for automated three-dimensional object reconstruction and inspection," *ACM Computing Surveys*, vol. 35, no. 1, pp. 64–96, Mar. 2003.

[8] S. D. Roy, S. Chaudhury, and S. Banerjee, "Active recognition through next view planning: A survey," *Pattern Recognit.*, vol. 37, no. 3, pp. 429–446, 2004.

[9] P. Quin, G. Paul, A. Alempijevic, D. Liu, and G. Dissanayake, "Efficient neighbourhood-based information gain approach for exploration of complex 3D environments," in *Proc. IEEE Int. Conf. Robotics Autom.*, Karlsruhe, May 2013, pp. 1343–1348.

[10] S. Kriegel, T. Bodenmüller, M. Suppa, and G. Hirzinger, "A surface-based next-best-view approach for automated 3D model completion of unknown objects," in *Proc. IEEE Int. Conf. Robotics Autom.*, Shanghai, May 2011, pp. 4869–7874.

[11] S. Foix, G. Alenyà, J. Andrade-Cetto, and C. Torras, "Object modeling using a ToF camera under an uncertainty reduction approach," in *Proc. IEEE Int. Conf. Robotics Autom.*, Anchorage, May 2010, pp. 1306–1312.

[12] Y. Li and Z. Liu, "Information entropy-based viewpoint planning for 3-D object reconstruction," *IEEE Trans. Robotics*, vol. 21, no. 3, pp. 324–337, 2005.

[13] S. Y. Chen and Y. F. Li, "Vision sensor planning for 3-D model acquisition," *IEEE Trans. Syst., Man, Cybern. B*, vol. 35, no. 5, pp. 894–904, Oct. 2005.

[14] C. Potthast and G. Sukhatme, "A probabilistic framework for next best view estimation in a cluttered environment," *J. Vis. Commun. Image Represent.*, vol. 25, no. 1, pp. 148–164, Jan. 2014.

[15] S. Foix, S. Kriegel, S. Fuchs, G. Alenyà, and C. Torras, "Information-gain view planning for free-form object reconstruction with a 3D ToF camera," in *Advanced Concepts for Intelligent Vision Systems*, ser. Lect. Notes Comput. Sci., vol. 7517, Brno, Sep. 2012, pp. 36–47.

[16] S. Foix, G. Alenyà, and C. Torras, "Lock-in time-of-flight (ToF) cameras: A survey," *IEEE Sensors J.*, vol. 11, no. 9, pp. 1917–1926, Sep. 2011.

[17] D. V. Lu, D. Hershberger, and W. D. Smart, "Layered costmaps for context-sensitive navigation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Chicago, Sep. 2014, pp. 709–715.

[18] L. Mihaylova, T. Lefebvre, H. Bruyninckx, K. Gadeyne, and J. D. Schutter, "A comparison of decision making criteria and optimization methods for active robotic sensing," in *Numerical Methods and Applications*, ser. Lecture Notes in Computer Science, I. Dimov, I. Lirkov, S. Margenov, and Z. Zlatev, Eds., vol. 2542. Springer, Jan. 2003, pp. 316–324.