

Task-driven active sensing framework applied to leaf probing

Sergi Foix^{a,*}, Guillem Alenyà^a, Carme Torras^a

^a*Institut de Robòtica i Informàtica Industrial, CSIC-UPC
Llorens i Artigas 4-6, 08028 Barcelona, Spain
Tel.: +34-93-4015751, Fax: +34-93-4015750*

Abstract

This article presents a new method for actively exploring a 3D workspace with the aim of localizing relevant regions for a given task. Our method encodes the exploration route in a multi-layer occupancy grid map. This map, together with a multiple-view estimator and a maximum-information-gain gathering approach, incrementally provide a better understanding of the scene until reaching the task termination criterion. This approach is designed to be applicable to any task entailing 3D object exploration where some previous knowledge of its approximate shape is available. Its suitability is demonstrated here for a leaf probing task using an eye-in-hand arm configuration in the context of a phenotyping application (leaf probing).

Keywords: Active Perception, Next Best View, Information Gain, Search Space Reduction

1. Introduction

The goal of task-driven exploration is to iteratively change the point of view so as to maximise the acquisition of information for solving a given task. We propose an algorithm that uses an information-gain criterion to compute the expected benefit of a set of candidate views, and combines it with other aspects,

*Corresponding author

Email addresses: sfoix@iri.upc.edu (Sergi Foix), galenya@iri.upc.edu (Guillem Alenyà), torras@iri.upc.edu (Carme Torras)

such as the proximity to the current view, to obtain the best next possible view at each iteration. Observe that this is a local approach, and that it cannot be globally optimal since each new position of the sensor only depends on the available information at each iteration. An optimal solution would require a
10 complete and accurate model beforehand.

This article emphasises the following four main ideas:

1. A multi-layer occupancy map approach can naturally encode all the knowledge: each layer codifies relevant information that is semantically different (Section 3.3). Particularly for leaf probing tasks, the space occupied
15 by the leaf and its surrounding clearance for allowing the tool to reach the leaf.
2. The importance of the termination criterion. Our representation, that explicitly represents the termination conditions in a specific occupancy layer, facilitates its definition and evaluation (Section 3.3).
- 20 3. Given a set of candidate viewpoints (Section 3.4), they can be evaluated using Information Gain (\mathcal{IG}), which can be easily defined and computed from the multi-layer representation (Section 3.5). A novelty in our proposal is that free space is also used for the \mathcal{IG} computation.
4. The accurate characterization of the sensor used, in our case a time-of-flight camera (ToF) (Section 3.6), plays an important role in the definition
25 and the computation of the \mathcal{IG} .

This article is an extended version of work published in [1]. We generalize our previous work by showing how the exploration models can be created depending on the characteristics of different tasks. Moreover, we demonstrate
30 through validation experiments the robustness of our multi-layer \mathcal{IG} criterion in conjunction with our frustum-based inverse sensor model, adding deeper explanations about the way to estimate the initial leaf pose.

2. Antecedents

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 large 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 [2], or on ground robots [3]. Obviously, climate conditions cannot 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 [4] (see Fig. 1). The throughput obtained in such installations is considerably high. However, sensors are located in a predefined position, mainly a top-view, that limits the possible photogrammetry tasks to be carried out, such as measuring the leaf length and rosette area [5], and extracting the plant 3D structure, either by using a depth camera [6], a stereo configuration [7], or a single RGB camera moving in a fixed direction [8]. Note that these systems can not deal with hard occlusions. Another main limitation 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 [9].

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 [10, 11]. In the context of the European project GARNICS³, an active perception system was proposed to overcome this weakness that involves ToF camera and a probing tool mounted on the end-effector of a robot

¹Image extracted from <http://www.phenome-fppn.fr/>

²Image extracted from <http://www.lemnatec.com/>

³<http://www.garnics.eu/>

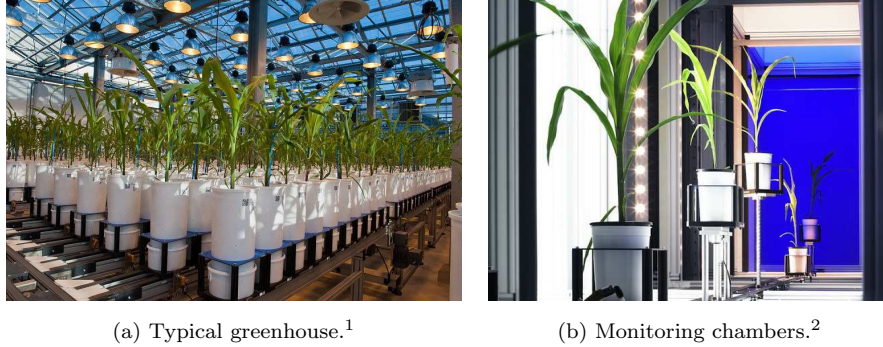


Figure 1: Example of a modern plant phenotyping greenhouse. (a) Plants are kept into labelled pots over conveyor belts to easily monitor them when being transported from the greenhouse to the watering, nutrient delivery or monitoring chambers. (b) Plants get measured one by one in the different monitoring chambers.

60 manipulator [12].

Classically, the *task* (usually implicit) in sensor path planning has been precise 3D object’s surface reconstruction [13, 14] and also object recognition [15]. Less commonly, sensor path planning has also been used to optimally segment particular object characteristics [16, 17] and to exploit sensor features for simplifying occlusions detection, formerly using a laser [18] and more recently using
 65 a ToF sensor [19].

Information gain has widely been used as *viewpoint selection criterion* in classical 3D modelling [20, 21]. Newer approaches compute the expected information gain using a discrete approximation of the sensor’s field of view based
 70 on pixel’s ray-tracing techniques. They can be single-resolution [22] or multi-resolution [23]. In contrast, our proposal uses the complete pixel’s frustum for a more accurate computation. Although a little bit slower, our approach assures to not miss any space between rays independently of the resolution of the octree. This may be negligible in large indoor/outdoor mapping [24], but it is crucial
 75 in our short range application.

An important peculiarity of our approach is that, since ToF sensors uncertainty is not uniform, a precise calibration of the sensor is required to adequately

model the image acquisition process and thus improve the \mathcal{IG} computation. In previous approaches, the sensor’s uncertainty is considered uniform for all the
80 acquired points. Up to our knowledge, the different uncertainty present in each pixel has not been used before for the precise computation of the \mathcal{IG} . Our model assures a good approximation of the sensor’s uncertainty distribution based on both pixel’s location and measured distance.

The most common way for representing the 3D space is by using occupancy
85 grid maps [24, 25]. An interesting idea for 2D robot navigation was proposed by Lu *et al.* [26], where a multilayer costmap approach was used to separate different semantic information on different layers. In this article, we extend this idea to 3D and proposed a new multi-layered occupancy map to separate the state map, the obstacle avoidance map and the termination criterion.

90 Finally, it is crucial to define a *termination criterion* that is relevant to the task. The most used ones have been the ratio of visible and occluded information in the scene [27], and different measures of the completeness of the model [14, 28]. However, completeness is usually non-intuitive to define. Vasquez *et al.* [25] proposed also to take into account if a mobile robot could
95 not find a path for any of the remaining candidate views. In contrast, in this work we link the task’s goal and the termination criterion by explicitly defining specific regions of interest (ROIs), one or more, in a unique 3D layer map.

3. Leaf probing

The main idea behind our method is that, based on the previous knowledge
100 of an exploration task, we can pre-establish a 3D occupancy map and a set of candidate views that, all together, indirectly serve as a guide to a Next-Best-View (NBV) planner for solving the task (Fig. 3). On the one hand, the 3D occupancy map is used for locating a set of possible regions of interest for the given task. On the other hand, the set of candidate views is used for reducing
105 the dimensionality of the gaze space while ensuring a complete coverage of the regions of interest.

The approach includes three main steps (see Fig.4):

1. selection of the target leaf from a plant
2. exploration of this leaf to gather enough relevant information for the task
110 at hand
3. effective execution of the task

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, is been carried out as proposed in Alenyà *et*
115 *al.*[9]. This article focuses on a method for solving the second step.

In the following sections we will explain every module of our approach in detail. Due to its high complexity, the task of leaf probing is the one used for illustration purposes and real experimental evaluation. Observe how this task does not require to have a complete leaf's model to accomplish its goal. Instead,
120 only specific regions in the leaf's contour are needed (Fig. 3f). 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 (Section 3.3).

3.1. Set-up

The experimental set-up consists of a Barrett WAM arm (robot manipulator)
125 and three sensors: a PMD Camboard (ToF camera), a SPAD meter (chlorophyll measuring tool), and a Kinect camera (see Fig. 2a). 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
130 selected target leaves (see Fig. 2b). 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.

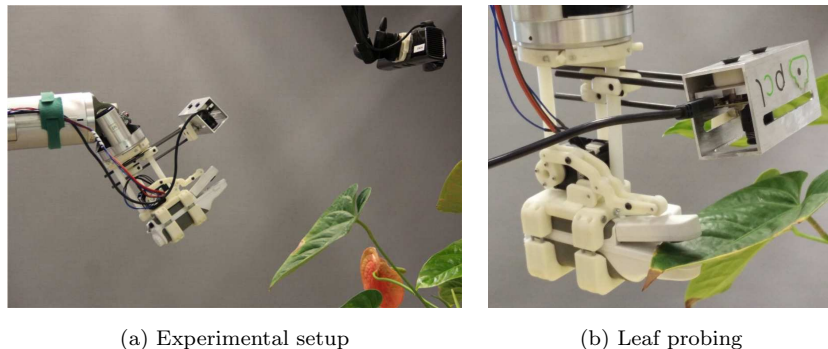


Figure 2: Snapshots of the leaf probing task: (a) Overall view of the complete setup, with the robot carrying the ToF camera and the tool, and the external Kinect camera. (b) Detail of the action of probing. Observe that the leaf probing task requires clearance (above and below) of a sector of the leaf.

3.2. Leaf pose estimation

135 The first demand within the system is to accurately localize a leaf in the scene (Fig. 4). As it will be seen in Section 3.6, correct depth measurements are only possible within the camera’s calibrated depth range. Thanks to a simple mean squared error (MSE) reduction approach we can guarantee a good leaf pose estimation (see Alg. 1). The camera is continuously repositioned to the
 140 same measure-based estimated pose, 35 cm. orthonormal to the target leaf, until the MSE between the camera’s current and desired locations reaches a minimum threshold.⁴ Leaf pose is acquired by computing the principal components of the depth measurements using the automatic leaf-extraction approach [9].

Once a leaf has been selected and its pose correctly estimated, an initial
 145 task-driven *exploration model* is defined. Its aim is to encode a set of multiple possible paths that allow the NBV planner to effectively fulfill the task (i.e. leaf probing). The exploration model is composed of a *multi-layer occupancy grid map* (Fig. 3f) and a *set of candidate viewpoints* (Fig. 3i). We recall that,

⁴As a practical note, the threshold must be chosen wisely to allow completion: just a little bit higher than the leaf’s pose estimation error, 1 mm. in our case.

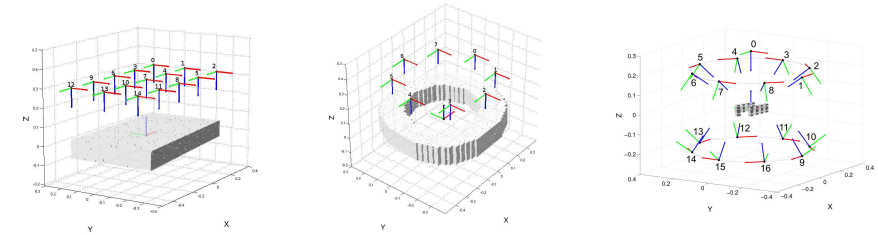
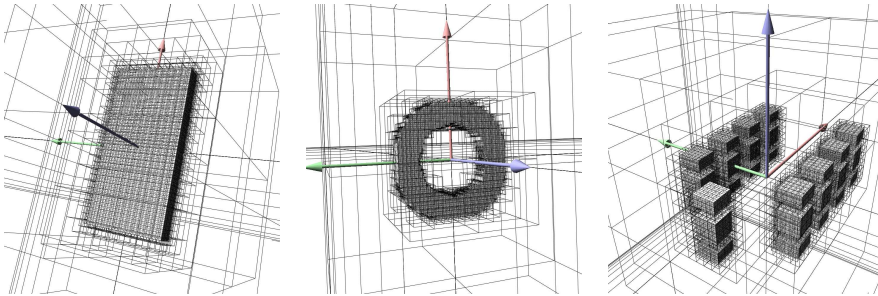
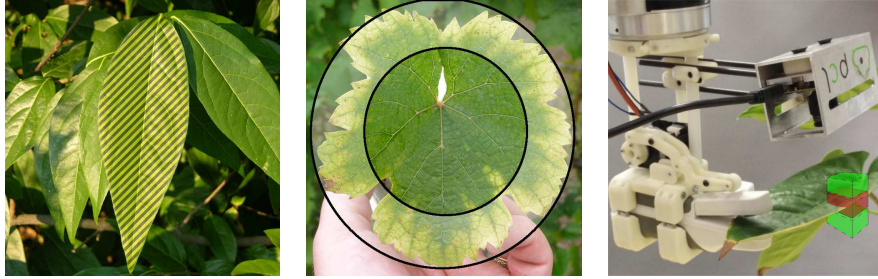


Figure 3: Example of plant monitoring tasks. This figure illustrates three clear examples of plant monitoring tasks (a-b-c), their corresponding expected occupancy grid maps (d-e-f) and their set of candidate views (g-h-i). Notice the strong correlation between task, occupancy maps and candidate views. Observe how, depending on the task, the expected occupancy map can be represented, with greater or less precision, by global (d) or by partial models (e, f). Also, pay attention to the variation in the number and location of the candidate views according to the task and its expected occupancy model. Notice how the leaf probing task needs bottom-up views to ensure a good coverage of the expected clearance under the leaf.

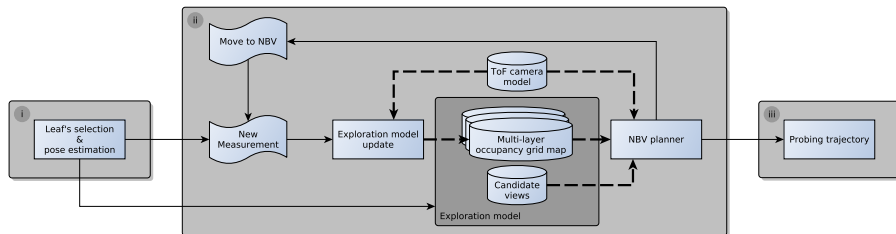


Figure 4: Overall block diagram of the leaf probing sensor planning framework.

although we concentrate on spatial restrictions like clearance, it is also shown
 150 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.
 Additionally, it is important to observe how our approach can be easily adapted
 to any kind of targeted object, no matter its size or shape. This is possible
 thanks to both, our highly reconfigurable multi-layer occupancy grid map and
 155 the unrestricted distribution of candidate views, as it will be seen in the following
 sections.

3.3. Multi-layer occupancy grid map

We have extended the idea to 3D occupancy maps, and propose to subdivide
 the occupancy representation of the exploration model into three semantically-
 160 separated layers $\{m^{\text{task}}, m^{\text{state}}, m^{\text{obs}}\}$. By doing this, the method obtains a
 wide versatility that facilitates four key aspects: the specification of the task
 termination criterion, the precise adjustment of particular exploratory attrac-
 tors, the correct treatment of the possible occluded regions of interest and the
 computation of obstacle avoidance trajectories.

165 The aim of the *task termination* layer m^{task} is twofold: first, to indicate
 whether the exploration can already be halted; and second, to act as a prior
 for the NBV planner. This layer is composed of what we call regions of interest
 (ROIs). Each ROI is defined as a region of expected occupancy in the model,
 and acts, by itself, as a global termination criterion. That means that if the
 170 expected occupancies within a certain ROI are fulfilled after a measurement,

Algorithm 1 Approximation Loop

APPROXIMATIONLOOP(\mathbf{x}_0, thr)

INPUTS:

 \mathbf{x}_0 : Initial sensor pose in global coordinates thr : Error threshold

```
1:  $i = 0$ 
2: while (  $MSE \geq thr$  ) do
3:    $\mathcal{D}_i \leftarrow \text{DATAACQUISITION}(\mathbf{x}_i)$ 
4:    $\mathbf{y}_i \leftarrow \text{LEAFPOSECOMPUTATION}(\mathcal{D}_i)$ 
5:    $\mathbf{x}_{i+1} \leftarrow \text{ORTHONORMALVIEW}(\mathbf{y}_i)$ 
6:    $MSE \leftarrow \text{MEANSQUAREERROR}(\mathbf{x}_i, \mathbf{x}_{i+1})$ 
7:    $\text{MOVEROBOT}(\mathbf{x}_{i+1})$ 
8:    $i = i + 1$ 
9: end while
```

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. 5b). Each ROI is composed of three bounding boxes or here called *bricks*, two of them labelled as free, one
175 at the top and one at the bottom, and another one labelled as occupied in the middle (Fig. 5c). 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
180 probing task.

The *state* layer m^{state} is the one keeping the complete update of all measurements taken during an experiment. As a result, and in conjunction with the m^{task} layer, the NBV planner can thereafter predict a more realistic estimation of the information gain. Such prediction is accomplished by simulation,
185 i.e., every candidate viewpoint is ray-traced over m^{state} . Once simulated, each new virtual measurement is updated into a copy \hat{m}_i^{task} of the global m^{task} layer

(Fig. 5a), for posterior computation of its expected \mathcal{IG} (Section 3.5). Observe that, although the cost of ray-tracing is very high, each view simulation is a completely independent process and, consequently, all possible candidate views
 190 can be computed in parallel at the same time.

The *obstacle avoidance* layer m^{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. As has been introduced before, other task-relevant components can be also included in a new map following the same
 195 approach. For example, the veins and the yellow spots on the leaves must be avoided when probing.

3.4. Candidate viewpoints

A predefined set of vantage points \mathcal{C} , such as:

$$\mathcal{C} = \{ \mathbf{cv}_i \mid i = 1, \dots, k \}, \quad (1)$$

must be chosen in order to guarantee a good coverage for all ROIs within the m^{task} layer (Fig. 3, last row). Candidate viewpoints can be classified as *re-*
 200 *dundant*, if they point to regions where some other viewpoints already do, or as *non-redundant*, if they are the only ones. Depending on the task, one type or another will be more or less propitious. For instance, if the task is to explore a big continuous area where occlusions can easily arise, then *redundant* viewpoints will be mandatory in order to guarantee the correct fulfillment of
 205 the task. On the other hand, tasks where the solution can be found on many different ROIs, such as the leaf probing, multiple *non-redundant* views are preferred.⁵ The final subset of viewpoints that will be chosen will depend on each scenario. Albeit locally predefined offline, both the *multi-layer occupancy grid map* and the set of *candidate viewpoints* are updated online. In our application,
 210 we assume complete confidence on the robot’s pose, and therefore *all candidate*

⁵Although the process of viewpoint simulation is parallelized, it is always recommended to have as less views as possible to reduce the computational load.

viewpoints are updated just once, based on the leaf pose estimation (initial state of the system). Note that it will be straightforward to take into account the uncertainty of the robot within the system by simply incorporating the leaf’s pose estimation module into the main loop and adapt the views accordingly.

215 Prior to the selection of the NBV and after the last measurement update, the system checks within the m^{task} layer for active 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 termination criterion,
 220 the planner will compute the next-best-view cv^* based on the \mathcal{IG} cost function.

3.5. NBV planner - Information gain

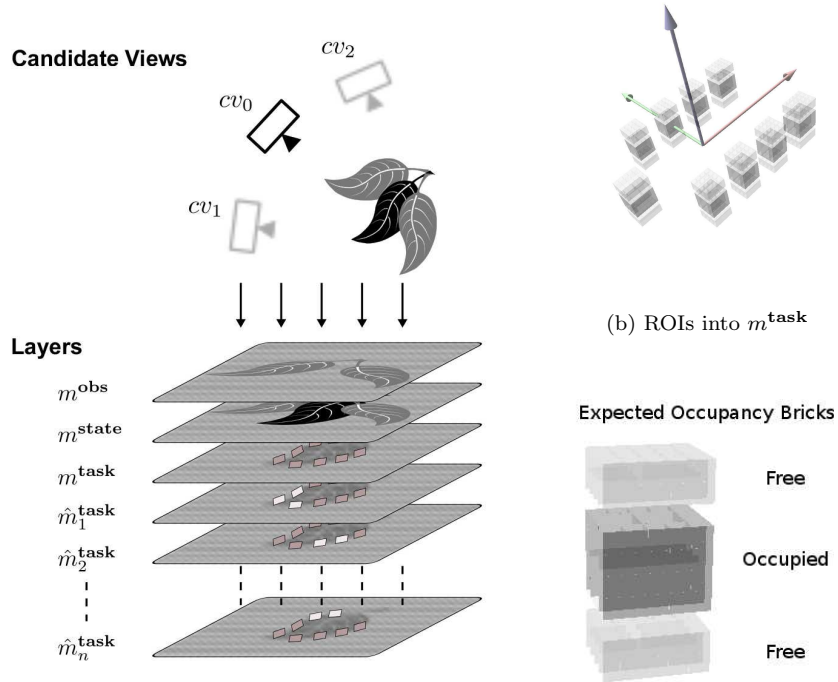
Our approach is based on the multi-objective performance criterion described by Mihaylova *et al.* in [29], where the final candidate view cv^* is selected not only by choosing the view with the highest \mathcal{IG} , but also by taking into account the travel distance to the rest of views:

$$cv^* = \arg \min_{\mathcal{J}} \|\mathcal{J}\|, \quad (2)$$

where $\|\cdot\|$ is the euclidean distance from the current view and \mathcal{J} is the set of views with expected information gains around the maximum, such as:

$$\mathcal{J} = \left\{ cv_j \mid \mathcal{IG}(cv_j) \geq \max(\mathcal{IG}(\mathcal{C})) - \delta \cdot \max(\mathcal{IG}(\mathcal{C})), j = 1, \dots, k \right\}, \quad (3)$$

where $\delta \in [0, 1]$ is the factor controlling the minimum \mathcal{IG} threshold. In our experiments we set it at 0.15. This way, closer views are preferred, even if they do not yield the highest \mathcal{IG} . As a consequence, we take advantage of intermediate
 225 views while also avoiding big jumps within the exploration trajectory. Figure 9 shows the behavior of the NBV planner when changing the value of δ . Although a more robot-specific travel cost distance could have been taken into account, e. g. the sum of the absolute difference in each pose’s joint angle [30], we decided to be as agnostic as possible to the type of robot.



(a) Schematic view of the multi-layer occupancy grid map.

(c) Bricks into the ROI.

Figure 5: Multi-layer occupancy grid map, ROIs and Bricks. (a) Each layer is a 3D occupancy map that semantically separates the different types of information. The potential information-gain obtained when visiting each candidate view cv_i is estimated from this multi-layer grid map. Each estimation \hat{m}_i^{task} is accomplished by frustum ray-tracing over the current map state m^{state} . (b) Graphical representation of the 9 ROIs within the task termination layer m^{task} . (c) Inner composition of *bricks* into a single ROI.

The expected information gain of each candidate viewpoint is computed as follows:

$$\mathcal{IG} = \sum_o \Delta H(\hat{m}_{i,o}^{\text{task}}) + \beta \sum_f \Delta H(\hat{m}_{i,f}^{\text{task}}), \quad (4)$$

230 where $\Delta H(\hat{m}_{i,o}^{\text{task}})$ and $\Delta H(\hat{m}_{i,f}^{\text{task}})$ are, respectively, the *expected information gains* of occupied and free bricks when going from the current m^{task} to the simulated \hat{m}_i^{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
 235 parameter is set to 1. Note that Eq. 4 can be also interpreted as a weighted *change of entropy* between prior and posterior probability density functions.

3.6. ToF camera model and map update

It is important to highlight the relevance of having a good characterization of the sensor’s depth model, like the one presented in Foix *et al.* [31]. As it
 240 has been seen in the previous sections, the sensor model is not just used for updating the occupancy grid maps but also for computing the information gain estimation when simulating the candidate views. 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.
 245 After calibration and filtering, the uncertainty of the remaining errors can be approximated by Gaussian noise with zero mean and uniform standard deviation and considered independent of any other measurement. Therefore, our model provides the uncertainty depending on the pixel location and the depth measurement received. The calibration process is long, tedious, and has to be
 250 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. In our experiments, based on the ToF camera’s field of view and the mean leaf size, we have selected a preferred distance of 35 ± 5 cm.

After a given depth measurement z_t at pose x_t , the expected occupancy, of
 255 every grid map cell \mathbf{c}_i within the calibrated range is computed as a combined

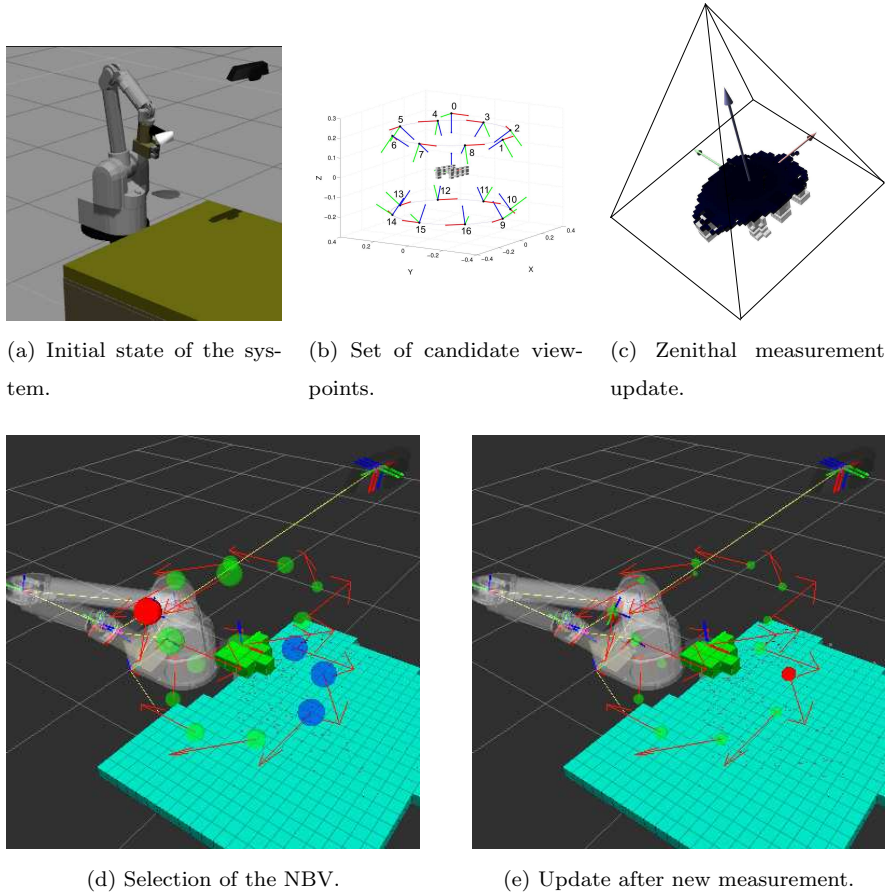


Figure 6: Step-by-step graphical interpretation of the task-driven active sensing framework approach. (a) The system assumes having a leaf in the field-of-view of the camera. Once the leaf is detected, the exploration model is introduced into the system (b), together with the first measurement (c). Views 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).

probability $p(\mathbf{c}_i|z_t, x_t)$ by multiplying both pixel-based (u, v) and depth-based $|z_t|$ independent probabilities such as:

$$p(\mathbf{c}_i|z_t, x_t) = \begin{cases} 0 & , \text{ if } |z_t| < \text{min_range} \\ p(\mathbf{c}_i|u, v, x_t) * p(\mathbf{c}_i||z_t|) & , \text{ if frustum in range} \cdot \\ 0 & , \text{ if } |z_t| > \text{max_range} \end{cases} \quad (5)$$

This combined probability is the one that determines the *inverse_sensor_model*. As defined in [32], the expected occupancy, or belief, is kept in the classical log odds form and fused with the following variation of the binary Bayes filter:

$$l_{t,i} = \max(\min(l_{t-1,i} + \text{inverse_sensor_model}(\mathbf{c}_i, x_t, z_t), l_{max}), l_{min}), \quad (6)$$

where $l_{t,i}$ is the posterior expected occupancy cell state, l_{min} and l_{max} define the boundaries of the belief cell state and

$$\text{inverse_sensor_model}(\mathbf{c}_i, x_t, z_t) = \log \left(\frac{p(\mathbf{c}_i|z_t, x_t)}{1 - p(\mathbf{c}_i|z_t, x_t)} \right). \quad (7)$$

Previous to any measurement, each expected occupancy cell is initialized with the state of unknown:

$$p(\mathbf{c}_i) = 0.5 \rightarrow l_{0,i} = 0. \quad (8)$$

3.7. Validation

Once the algorithm is running, it becomes quite complicated to see why the algorithm considers certain candidate views and not others. As a consequence, it is difficult to check whether the approach is working as expected or not. With

the aim of better understanding each NBV selection, we have incorporated a special visual marker into our monitoring interface. This marker, that is based on a sphere, reveals two things. On the one hand, its colour highlights the current state of the candidate view: selected (**red**), not selected (**green**) or unreachable (**blue**). And, on the other hand, the length of its radius indicates its expected \mathcal{IG} s (Fig. 6).

One of the core parts of our approach is the proper estimation of each candidate view \mathcal{IG} through simulation. As it has been already introduced in Section 3.3, such simulation is carried out by means of complete⁶ frustum ray-tracing over the current map state m^{state} (Fig. 7a). As a "proof of concept", we have performed a simple experiment in order to validate and demonstrate the functionality of our approach. The experiment consists in evaluating, over a simple planar multi-layer occupancy grid model, a wide range of views at multiple orientations and distances, see Fig. 7b. Observe how views at the intermediate semicircle, at 35 cm. far, produce greater \mathcal{IG} estimations than views that are closer or farther. A proof of how well integrated is the ToF camera model. It can also be clearly seen how \mathcal{IG} gets reduced around lateral views.

Now that the reader is more familiar with the computation of the candidate views \mathcal{IG} and its symbolic representation, it is time to observe how the algorithm behaves in a more realistic scenario; for instance, the exploration task required for computing the area of a leaf, previously illustrated in Fig. 3. Two different setups are considered for this evaluation, one with a clear view of the leaf (Fig. 8a) and another one with an obstacle partially occluding the leaf (Fig. 8b). A very special candidate view has been added into the system in order to show how important certain views can be for taking into account possible contingencies, such as occlusions, in the scene. The set of candidate views consists of 25 fronto-parallel views and an extra one, intentionally located at the right side of the scene with a 45° tilt, in such a way that it provides a clear observation of the

⁶Each pixel's frustum is computed, instead of simple ray-tracing. This ensures a more accurate simulation.

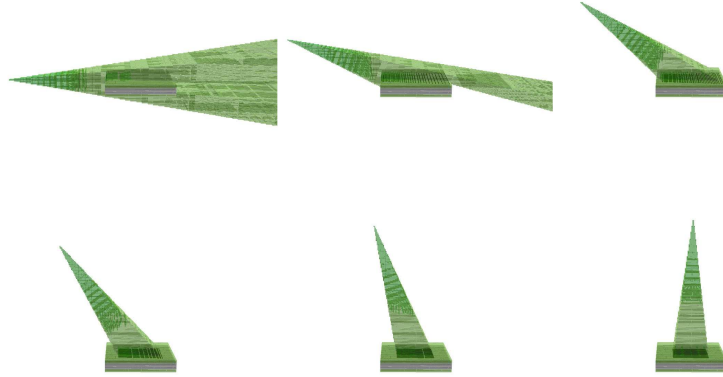
295 target leaf when occluded. Observe how in the absence of obstacles (Fig. 8a),
 and after the initial view, none of the peripheral views, including the special
 extra lateral view, provides any significant increment of \mathcal{IG} . But if we deliber-
 ately occlude the left part of the leaf, a very different expected \mathcal{IG} distribution
 is observed after the initial view (Fig. 8b). This time, as expected, that extra
 300 lateral view is the one providing the maximum \mathcal{IG} , and therefore it is correctly
 selected. Observe that, once the targeted leaf is correctly localized, the system
 does not need to explicitly to detect the obstacles since it deals with them by
 incorporating the measurements into the corresponding model layer. It is also
 important to notice that, in both cases, the selected NBVs are those with the
 305 maximum expected \mathcal{IG} . Such behavior is the consequence of setting $\delta = 0.0$.
 Figure 9 clarifies the behavior of the NBV planner when δ is differently tuned
 (see Eq. 3). It can be seen that, as long as a rich set of candidate views is pro-
 vided, smoother final trajectories can be obtained by increasing its value. We
 remark that if δ gets too high the algorithm will consider all views as possible
 310 candidates and may, therefore, fall into an infinite loop of refinements, similar
 to a local minimum. By banning already visited candidates, this situation is
 easily avoided.

4. Experiments

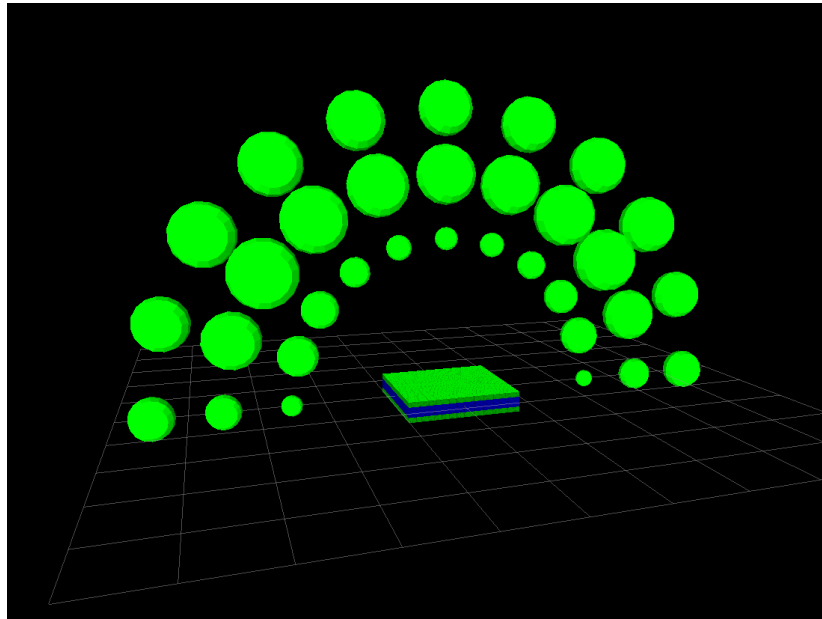
Experiments have been carried out under simulation and subsequently tested
 315 on the real robot, in a very similar scenario using the very same algorithm and
 parametrization (Figs. 10 and 11, respectively).

4.1. Simulation of leaf probing

This experiment analyses the behavior of the system in an isolated leaf and
 multiple-leaves scenes, Fig. 10a and 10b respectively. The goal is to first eval-
 320 uate the suitability of the method on both tidy and cluttered scenes, and then
 compare the robot behavior between simulation and reality in the following sec-
 tion. Figures 10c and 10d display the priorities on the selection of views while

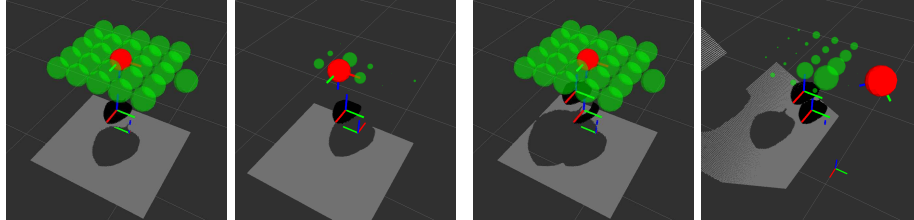


(a) Graphical interpretation of the frustums in a subset of simulated candidate views.



(b) Graphical interpretation of the \mathcal{IG} s distribution at 25, 35 and 45 cm.

Figure 7: Experiments for validating the \mathcal{IG} distribution over candidate views. The validation of the approach consists in evaluating the \mathcal{IG} distribution around a planar occupancy grid model. Each candidate view is simulated by computing the complete frustum of the sensor (a). Three sets of views, at different distances around a semicircle, are evaluated (b). Observe how, the value of \mathcal{IG} varies not only with the number of occupied and free cells observed within the sensor's field of view, but depending on the depth of the measurements, thanks to the accurate characterization of the sensor after calibration.

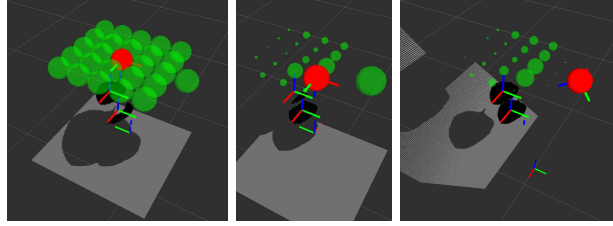


(a) Scene without occlusions ($\delta = 0.0$). (b) Scene with a partial occlusion ($\delta = 0.0$).

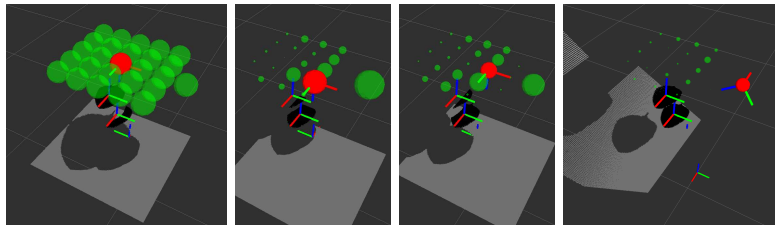
Figure 8: Simulated experiments for computing the area of a leaf. Two particular scenarios: (a) an ideal one without any obstacle blocking the target leaf, and (b) in the presence of a partial occlusion on one side of the leaf. In both experiments, after an initial fronto-parallel measurement, at 35 cm. above the center of the leaf, a single NBV is required before considering the exploration task as concluded. Observe that in both cases the selected NBVs are those with the maximum expected \mathcal{IG} , as a consequence of setting $\delta = 0.0$.

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
 325 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, so we can see how the system reacts. Observe, for instance, the behavior in the cluttered scenario: the planner chooses view 13 in
 330 the first place, as the closest view among of those with the highest \mathcal{IG} ; since it is unreachable, this view is removed from the candidate views list and following the same criterion a NBV is chosen again until a reachable view is selected. The final solution for both experiments is a top lateral-left view (5 or 6). Comparing both priority figures it can also be noticed how priorities change depending on
 335 the scene and that some of the views, which were taken into account in the isolated leaf experiment, have been substituted by other ones in the cluttered experiment due to the presence of obstacles; for instance view number 1 for view number 8.

Table 1 shows how the presence of obstacles also affects the number of NBVs
 340 required until achieving the termination task criterion. In the isolated leaf scene,



(a) Scene with a partial occlusion ($\delta = 0.15$). Intermediate views, with very high \mathcal{IG} , are now taken into account as possible NBVs. The closest view to the current pose is the one selected.



(b) Scene with a partial occlusion ($\delta = 0.25$). The higher the δ , the greater the number of selected intermediate views.

Figure 9: Simulated experiments illustrating the behaviour of the NBV planner when δ factor is being increased. The very same scenario from Fig. 8b is used.

most of the experiments (61%) fulfill the task termination conditions at the second view⁷. This is good news for plant phenotyping where high throughput is required. On the contrary, when obstacles are present, the task finishes with a second view 33% of the times, and requires at least one more view 56% of the times.

345

4.2. Real

We carried out a set of 20 real experiments, using the same algorithm and parametrization as in the simulation, and with very similar scenes, see Fig. 11. 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 simulations; it tried to go to the bottom and more informative views, and when

350

⁷The initial zenithal measurement is taken into account.

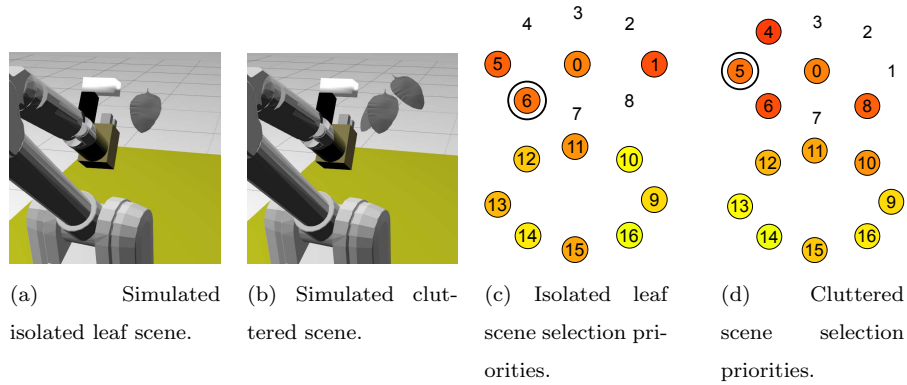


Figure 10: Isolated and cluttered scenes for the simulated experiments. Figures (c) and (d) show the priority of views in scenes (a) and (b), respectively. Each number refers to a candidate view, see Fig. 6b. Colors indicate priority, going from yellow, the highest, to red, the lowest. Take into account that priorities are computed after a first zenithal measurement of the leaf. That is the reason why highest priorities are observed at the bottom views. The most selected view for the final leaf probing is marked with an extra-circle. These results have been extracted after 100 simulations per scene.

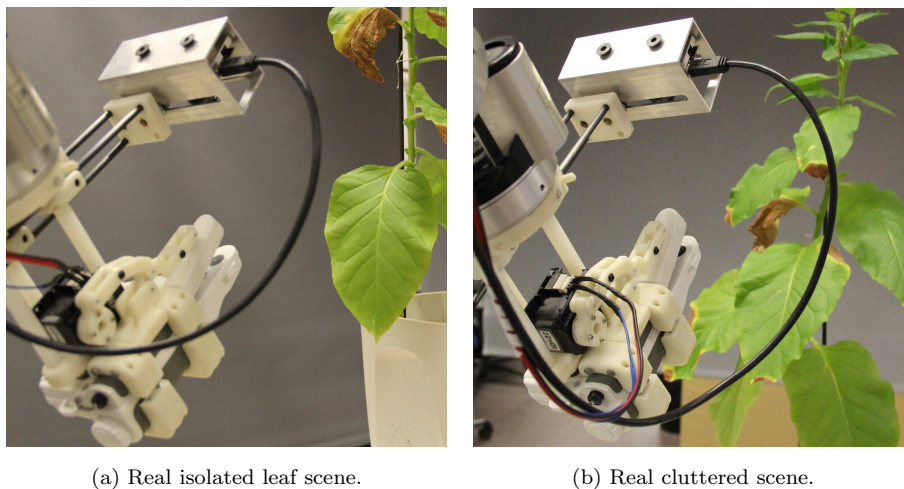


Figure 11: Isolated and cluttered scenes for the real experiments.

N. of views	\neg obstacle	obstacle
2	61%	33%
3	27%	56%
4	5%	2%
5	1%	0%
6	1%	0%
(-)	5%	9%

Notes: Some experiments resulted inconclusive (-) due to external errors in the path planner.

Table 1: Percentage of experiments that finished with respect to the number of views. 100 Experiments carried out in simulation.

not possible it chose the NBV until the fulfilment of the termination criterion, and the subsequent leaf probing action. 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 very close to each other and the difference in the obtained information gain is small.⁸

5. Conclusions

This article presented a complete task-driven active sensing framework and its particular suitability for complex plant phenotyping tasks such as autonomous leaf probing. Our solution is composed of a manipulator robot carrying a ToF camera and a specialized probing tool. Although our experiments include either a tool for chlorophyll measuring or a sampling tool for leaf probing, both tasks involve the same framework: the robot autonomously changes the point of view of the camera to take new images, and when enough information is gathered

⁸Additional material at:
<http://www.iri.upc.edu/groups/perception/leafProbing>.

365 the leaf probing action 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.

We have proposed a novel 3D task representation based on a multi-layer occupancy grid map, where each layer codifies different semantically relevant information. It has been shown that this representation has four main advantages. First, it allows to encode tasks based on prior knowledge using a combination of free and occupied space, and even possible constraints. Second, it facilitates the fusion of measurement uncertainties, thanks to both a Bayes filter and an accurate calibration of the ToF camera. Third, it has been shown that IG can be effectively used to select the next-best-view; a formulation has been introduced to compute the expected information gain from this representation. As the application may require to minimize the motion of the robot, a criterion has also been introduced to prefer closer views even if they provide slightly less information. And fourth, 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.

While the current work has taken for granted the generation of the set of candidate views, future work will focus on exploring how to automatically generate them such that the set will only contains the most *task-significant* views. Currently, we are working on doing it off-line through multiple simulations; in such a way that the possible contingencies can be included within the simulations, allowing us to extract, in the form of an expert rule system, the most informative sequence of views depending on such events. These task-based expert rule system will substitute the current proximity rule used for NBV selection when information-gain ambiguity arises.

390 Results on both simulated and real experiments validate the proposed approach for the leaf probing task and allow us to envision a good performance on other 3D object exploration tasks, provided some previous knowledge of its general shape is available. Experiments have also 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 con-

trol 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.

6. Acknowledgements

400 This research has been partially supported by the European project GARNICS FP7-ICT-2009-4-247947, CSIC project MANIPlus 20350E102 and by the Spanish State Research Agency through the María de Maeztu Seal of Excellence to IRI (MDM-2016-0656).

References

- 405 [1] S. Foix, G. Alenyà, C. Torras, 3D Sensor planning framework for leaf probing, in: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Hamburg, 2015, pp. 6501–6506.
- [2] I. Colomina, P. Molina, Unmanned aerial systems for photogrammetry and remote sensing: A review, *ISPRS Journal of Photogrammetry and Remote Sensing* 92 (0) (2014) 79 – 97.
- 410 [3] S. Nielsen, K. Jensen, A. Bøgild, O. J. Jørgensen, N. Jacobsen, C. L. Jæger-Hansen, R. N. Jørgensen, A low cost, modular robotics tool carrier for precision agriculture research, in: 11th Int. Conf. on Precision Agriculture, 2012.
- 415 [4] B. Berger, High-throughput phenotyping of cereals at the plant accelerator, in: Plant and Animal Genome XXI Conf., Plant and Animal Genome, 2013.
- [5] N. An, C. M. Palmer, R. L. Baker, R. C. Markelz, J. Ta, M. F. Covington, J. N. Maloof, S. M. Welch, C. Weinig, Plant high-throughput phenotyping using photogrammetry and imaging techniques to measure leaf length and rosette area, *Computers and Electronics in Agriculture* 127 (2016) 376–394.
- 420

- [6] Y. Chn, D. Rousseau, P. Lucidarme, J. Bertheloot, V. Caffier, P. Morel, tienne Belin, F. Chapeau-Blondeau, On the use of depth camera for 3d phenotyping of entire plants, *Computers and Electronics in Agriculture* 82 (2012) 122–127.
- 425 [7] E. E. Aksoy, A. Abramov, F. Wrgtter, H. Scharr, A. Fischbach, B. Dellen, Modeling leaf growth of rosette plants using infrared stereo image sequences, *Computers and Electronics in Agriculture* 110 (2015) 78–90.
- [8] S. Jay, G. Rabatel, X. Hadoux, D. Moura, N. Gorretta, In-field crop row phenotyping from 3d modeling performed using structure from motion, 430 *Computers and Electronics in Agriculture* 110 (2015) 70 – 77.
- [9] G. Alenyà, B. Dellen, S. Foix, C. Torras, Robotized plant probing: Leaf segmentation utilizing time-of-flight data, *IEEE Robotics Autom. Mag.* 20 (3) (2013) 50–59.
- [10] R. Bajcsy, Active perception, *Proc. IEEE* 76 (8) (1988) 996–1005.
- 435 [11] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, F. Nuflo, Modeling visual attention via selective tuning, *Artif. Intell.* 78 (1-2) (1995) 507–545.
- [12] G. Alenyà, S. Foix, C. Torras, Tof cameras for active vision in robotics, *Sensors and Actuators A: Physical* 218 (2014) 10–22.
- 440 [13] W. R. Scott, G. Roth, J. F. Rivest, View planning for automated three-dimensional object reconstruction and inspection, *ACM Computing Surveys* 35 (1) (2003) 64–96.
- [14] S. Kriegel, C. Rink, T. Bodenmüller, M. Suppa, Efficient next-best-scan planning for autonomous 3d surface reconstruction of unknown objects, 445 *Journal of Real-Time Image Processing* 10 (4) (2015) 611–631.
- [15] S. D. Roy, S. Chaudhury, S. Banerjee, Active recognition through next view planning: A survey, *Pattern Recognit.* 37 (3) (2004) 429–446.

- [16] C. Madsen, H. Christensen, A viewpoint planning strategy for determining true angles on polyhedral objects by camera alignment, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (2) (1997) 158–163.
- 450
- [17] A. Saxena, M. Sun, A. Y. Ng., Make3d: Depth perception from a single still image, in: *Proc. 23th AAAI Conf. Artif. Intell.*, Chicago, 2008, pp. 1571–1576.
- [18] J. Maver, R. Bajcsy, Occlusions as a guide for planning the next view, *IEEE Trans. Pattern Anal. Mach. Intell.* 15 (5) (1993) 417–433.
- 455
- [19] S. Foix, G. Alenyà, C. Torras, Towards plant monitoring through next best view, in: *Proc. 14th Int. Conf. Cat. Assoc. Artif. Intell.*, Lleida, 2011.
- [20] B. Curless, M. Levoy, A volumetric method for building complex models from range images, in: *Computer Graphics. Proc. ACM SIGGRAPH Conf.*, ACM Press, New York, 1996, pp. 303–312.
- 460
- [21] E. López-Damian, G. Etcheverry, L. E. Sucar, J. López-Estrada, Probabilistic view planner for 3D modelling indoor environments, in: *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, St. Louis, USA, 2009, pp. 4021–4026.
- [22] C. Potthast, G. Sukhatme, A probabilistic framework for next best view estimation in a cluttered environment, *J. Vis. Commun. Image Represent.* 25 (1) (2014) 148–164.
- 465
- [23] J. I. Vasquez-Gomez, L. E. Sucar, R. Murrieta-Cid, Hierarchical ray tracing for fast volumetric next-best-view planning, in: *Proc. Int. Conf. Comput. and Robot Vis.*, Regina, Canada, 2013, pp. 181–187.
- 470
- [24] F. Bissmarck, M. Svensson, G. Tolt, Efficient algorithms for next best view evaluation, in: *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Hamburg, Germany, 2015, pp. 5876–5883.

- 475 [25] J. I. Vasquez-Gomez, L. E. Sucar, R. Murrieta-Cid, View/state planning for three-dimensional object reconstruction under uncertainty, *Autonomous Robots* (2015) 1–21.
- [26] D. V. Lu, D. Hershberger, W. D. Smart, Layered costmaps for context-sensitive navigation, in: *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Chicago, 2014, pp. 709–715.
- 480 [27] J. E. Banta, L. M. Wong, C. Dumont, M. A. Abidi, A next-best-view system for autonomous 3-D object reconstruction, *IEEE Trans. Syst., Man, Cybern. A* 30 (5) (2000) 589–598.
- [28] L. Torabi, G. K., An autonomous six-dof eye-in-hand system for in situ 3d object modeling, *Int. J. Robotics Res.* 31 (1) (2012) 82–100.
- 485 [29] L. Mihaylova, T. Lefebvre, H. Bruyninckx, K. Gadeyne, J. D. Schutter, A comparison of decision making criteria and optimization methods for active robotic sensing, in: I. Dimov, I. Lirkov, S. Margenov, Z. Zlatev (Eds.), *Numerical Methods and Applications*, Vol. 2542 of *Lecture Notes in Computer Science*, Springer, 2003, pp. 316–324.
- 490 [30] P. Quin, G. Paul, A. Alempijevic, D. Liu, G. Dissanayake, Efficient neighbourhood-based information gain approach for exploration of complex 3D environments, in: *Proc. IEEE Int. Conf. Robotics Autom.*, Karlsruhe, 2013, pp. 1343–1348.
- [31] S. Foix, G. Alenyà, C. Torras, Lock-in time-of-flight (ToF) cameras: A
495 survey, *IEEE Sensors J.* 11 (9) (2011) 1917–1926.
- [32] S. Thrun, W. Burgard, D. Fox, *Probabilistic Robotics*, MIT Press, Cambridge, 2005.