

Information-Gain View Planning for Free-Form Object Reconstruction with a 3D ToF Camera

Sergi Foix¹, Simon Kriegel², Stefan Fuchs², Guillem Alenyà¹, and Carme Torras¹

¹ Institut de Robòtica i Informàtica Industrial, CSIC-UPC,
Llorens i Artigas 4-6, 08028 Barcelona, Spain,
sfoix, galenya, torras@iri.upc.edu

² Institute of Robotics and Mechatronics, German Aerospace Center (DLR),
82234 Oberpfaffenhofen, Germany,
simon.kriegel, stefan.fuchs@dlr.de

Abstract. Active view planning for gathering data from an unexplored 3D complex scenario is a hard and still open problem in the computer vision community. In this paper, we present a general task-oriented approach based on an information-gain maximization that easily deals with such a problem. Our approach consists of ranking a given set of possible actions, based on their task-related gains, and then executing the best-ranked action to move the required sensor.

An example of how our approach behaves is demonstrated by applying it over 3D raw data for real-time volume modelling of complex-shaped objects. Our setting includes a calibrated 3D time-of-flight (ToF) camera mounted on a 7 degrees of freedom (DoF) robotic arm. Noise in the sensor data acquisition, which is too often ignored, is here explicitly taken into account by computing an uncertainty matrix for each point, and refining this matrix each time the point is seen again. Results show that, by always choosing the most informative view, a complete model of a 3D free-form object is acquired and also that our method achieves a good compromise between speed and precision.

1 Introduction

Viewpoint planning tries to exploit the process of modifying the pose of a sensor to acquire a new view of the scene. All tasks requiring different views (modelling, recognition, feature discovery...) can be interpreted as an information gain process, since an increment of information is expected with every new view. This information is classically used for geometrical modelling, but should not be limited to it, and also may include other characteristics such as leaf contours for plant segmentation [6] or wrinkles on clothes for grasping [14]. Specially when dealing with unknown scenarios, the system should decide actions based only on the available information and the expected reward of executing the selected action. In such scenarios, the ability to explicitly measure the gain of each action is crucial, and it is closely related to the internal representation used.

Active view planning becomes a space characterization task whose goal is to answer the question: *where should the sensor be placed for locating specific characteristics?*

Because it involves spatial characteristics (or at least located in space), the proposed approach uses a voxelized space where each voxel contains a complete 3×3 covariance matrix. This representation allows to account not only for exploration (unknown areas) but also for refinement, that is, the information gain of seeing characteristics again from a different point of view.

In summary, this paper brings the following contributions: A) an algorithm to select the most informative action from a given set for general view planning tasks; B) a convenient representation of the informative characteristics and their position on the space using a 3×3 covariance matrix for each one, and an efficient implementation using a multiresolution octree; C) an efficient method to compute the expected gain of a new data acquisition fusing information from exploration and from refinement, that accounts explicitly for the orientation of the sensor and for the acquisition covariance matrix.

This paper is organized as follows: Sec. 2 summarizes related work; in Sec. 3 ToF sensors and their acquisition uncertainty are introduced; Sec. 4 presents the overall algorithm structure; the procedure used in the experiments to obtain the list of viewpoint candidates is presented in Sec. 5; Sec. 6 introduces the proposed view planning method based on information-gain, while Sec. 7 presents some results using real data; and finally, Sec. 8 discusses the main conclusions.

2 Related Work

Sensor view planning has been commonly used for the tasks of precise geometrical model construction and object recognition (see the reviews [18] and [15]), and to a lesser extent for the optimal segmentation of particular object characteristics [11,16] and to exploit sensor features to easily detect occlusions, formerly using a laser [12] and more recently with a ToF sensor [6]. These algorithms can be classified according to the constraints they impose, on the type of objects that can handle, the sensors they use, the restrictions of the sensor positioning system, and more important, the decision-making strategy and the symbolic object representation they used.

In [17] objects are represented statistically by multidimensional receptive field histograms, and the camera is controlled by making hypotheses on the salient points of the previously learned objects and then moving to the most discriminative viewpoint. In [3] reinforcement learning is used to associate the current state with camera actions and their corresponding reward. Here the model is a particle representation, and it is updated with new sensor readings with the Condensation algorithm. More recently, a boost-based algorithm to combine different appearance estimators [9] has been proposed to compute the next view in a rotating object framework.

All the previous algorithms require some degree of training. When training is not applicable or too expensive, approaches using information-gain measures are a good alternative. In such approaches, two steps can be distinguished: the generation of a set of viewpoint candidates and the ranking of such candidates by evaluating the expected information gain of each action. Again, for viewpoint generation, the internal representation of the environment model plays an important role. Surface-based methods provide a set of viewpoints based on the location of jump edges [13], the trend

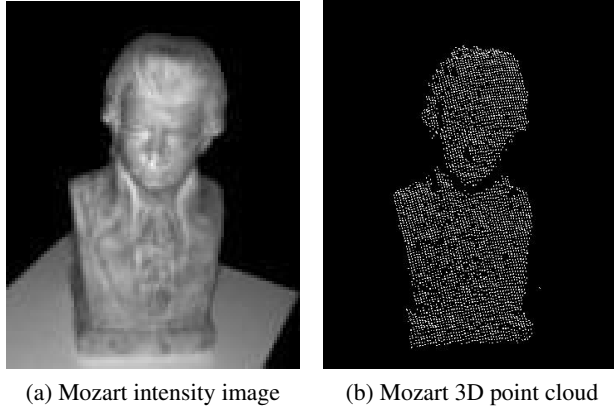


Fig. 1: Typical intensity image and matched point cloud acquired by a Swissranger SR4000.

of a contour [10] or the fitting of a parametric surface representation [1]. Volumetric methods provide viewpoints using the information of visited and non-visited portions of the workspace, and generally encode this space using voxel representations (or, more efficiently, octrees).

Information gain has been used before as viewpoint selection criterion in classical object modelling, where the sensor uncertainty is modelled using only the viewing direction and is considered uniform for all the acquired points. While some approaches require some degree of overlap to match consecutive sensor readings, other methods do not and consider this to be a positive feature (see [2] for a review). This is true for precise sensors, and for precise positioning systems, but it is not so when considering noisy sensors, specially when sensor readings have different uncertainties depending on their position on the image, as it is here the case with 3D ToF cameras.

This paper proposes a different approach for viewpoint evaluation. Independently of the viewpoint generation algorithm used, it relies on a volumetric space representation to encode the complete covariance matrix of the observed characteristics. The (possibly non-uniform) uncertainty in the acquisition process is explicitly used to compute the information gain produced by revisiting characteristics with large uncertainties (herein, overlapping is encouraged). This is combined with the information gain produced by exploring new areas to produce the total information gain for each different action.

3 3D ToF Camera

Depth measurements for the modelling task are captured by a 3D ToF camera. 3D ToF cameras have several advantageous features that make them more suitable for short-range applications than other range measurement technologies, such as laser scanners, stereo camera systems or the Kinect sensor. First, a 3D ToF camera provides registered depth and intensity images at frame-rates of up to 50 Hz (see Fig. 1a and 1b). Thus, they perform faster than sequentially measuring laser scanners. Second, a 3D

ToF camera features an illumination unit, which makes it independent from external light sources. Unlike stereo camera systems, a 3D ToF camera does not depend on textures. Finally, the minimal depth measuring range can get as close as 0.2 m³. This is necessary, because the sensor is attached to the tool-center point (TCP) of an articulated robot. There is thus only a small working range where the robot’s TCP can reach all the poses required to capture the desired object spherically. Consequently, the sensor has to approach the object as close as possible. In contrast, a Kinect sensor can only measure ranges from 0.6 m with the current specifications. The Kinect sensor is thus not suited for such modelling tasks.

However, 3D ToF camera data lacks accuracy and precision due to systematic errors and noise. There are two types of measurement noise: noise generated by the electronics and the so-called “jumping edges”. The former is limited by appropriately adjusting the integration time. The latter is caused by large depth-jumps within the scene, which are measured by the very same pixel. The resulting spurious measurement is handled by a specialized edge filter [13].

The systematic errors can be classified into intrinsic and extrinsic ones. The intrinsic systematic errors comprise the widely known distance- and amplitude-related errors. These errors are highly reduced by calibration procedures as proposed by [7]. The extrinsic systematic errors are generated by multiple light reception. On the one hand, there is light scattering, caused by imperfect optics and inter-reflections within the camera. This phenomenon is significant in case that there are large amplitude differences in the image, whereby the distance measurements with low amplitudes are affected by distance measurements with stronger amplitudes. The reasons are either large distances between foreground and background objects or large viewing angles. Both cases are negligible in our modelling task, because the desired object is always at the nearest distance in the sensor’s field of view and consequently the most illuminated. For a more detailed and broad classification and explanation of the different error sources, advantages and limitations of 3D ToF cameras, please refer to [5].

4 View Planning Procedure

The initial configuration of the system consists of placing the sensor at a certain distance in such a way that its field-of-view covers part of the scene and it remains in its focused range. Once the initial configuration is reached, the proposed method can be synthesized by four iterative steps (see some details in algorithm 1). First, data is acquired by means of a sensor (e.g. a 3D ToF camera). In the second step, two representations of the scene, one for the view generator and one for the information gain estimator (i.e. a 3D occupancy grid for active view planning), are updated with the new sensor measurements. During the third step, a set of candidate viewpoints is computed using the viewpoint generator. see Sec. 5. Finally, the view with the highest information gain is selected after simulating each candidate viewpoint, see Sec. 6.

³ Measures extracted with a Swissranger SR4000 camera decreasing its integration time to 1.0 ms.

Algorithm 1 Autonomous active view planning

 $\mathcal{M} \leftarrow \text{ACTIVEVIEWPLANNING}(\mathbf{x}_0, \Sigma_s, \mathcal{O}, \mathcal{S})$

INPUTS:

 \mathbf{x}_0 : Initial sensor pose in global coordinates. Σ_s : Sensor measurement covariance matrix. \mathcal{O} : 3D occupancy grid. \mathcal{S} : A set of n measurements.

OUTPUT:

 \mathcal{M} : Task-based representation.1: $i = 0$ 2: **repeat**3: $\mathcal{S}_i \leftarrow \text{DATAACQUISITION}(\mathbf{x}_i)$ 4: $(\mathcal{M}, \mathcal{O}) \leftarrow \text{UPDATEREPRESENTATIONS}(\mathcal{S})$ 5: $\mathbf{c}_m \leftarrow \text{VIEWPOINTSGENERATION}(\mathbf{x}_i, \mathcal{M})$ 6: $\mathbf{x}_{i+1} \leftarrow \text{DECISIONMAKER}(\mathbf{c}_m, \mathcal{O}, \Sigma_s)$ 7: $i = i + 1$ 8: **until** task completed9: **return** \mathcal{M}

5 Viewpoint Generation

In order to determine the following viewpoint accordingly to the information gain, a search space consisting of multiple viewpoints (possible sensor positions and orientations) is required as input. Since the workspace around an object features an infinite number of views, many authors reduce the search space by sampling candidate views around an approximate sphere or cylinder. Their candidate views always point to the center of their figures and, consequently, the sensor can not be positioned in a way that achieves optimal modelling results.

In this work, the *Viewpoint Estimator* [10] algorithm will be used. This algorithm that generates viewpoints by detecting boundary trends in a triangular mesh. It works as follows. Once new 3D data are acquired, a triangular mesh is reconstructed in a real-time stream. A quadratic patch is then fitted to each boundary region and new viewpoints, perpendicular to those patches, are then generated. Therefore, the search space is not limited to a set of pre-defined poses over a sphere or cylinder but it allows for any position and orientation. Depending on their position, relative to the sensor, the detected boundaries are classified as left, right, top and bottom. In their work, the next viewpoint was chosen heuristically by first going through the left, then right, top and bottom boundaries. Figure 2 shows an example of two boundaries classified as left and the subsequent region growing, which is used to fit a quadratic patch.

6 View Planning: Information Gain Ranking Criterion

In Information Theory, *information gain* is a probabilistic measure of how significant a new state estimate of the environment is. The concept of information gain is equivalent

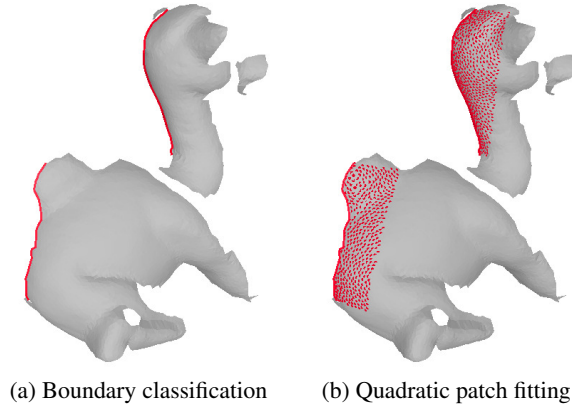


Fig. 2: Example of two boundaries obtained from a partial camel mesh, which are classified as left boundaries. A region growing is performed in order to fit a quadratic patch.

to the one of uncertainty or entropy reduction. Entropy, as defined by [19], is computed as:

$$H(x) = - \sum_X p(x) \log p(x), \quad (1)$$

where X is a finite set of values of a discrete random variable x that has $p(x)$ as probability distribution function. For a n multivariate Gaussian distribution with covariance matrix Σ , entropy can be computed as:

$$H(x) = \frac{1}{2} \log((2\pi)^n |\Sigma|). \quad (2)$$

As [20] already pointed out, using the determinant over all possible measurements for computing the information gain is computationally expensive. Based on his work, our approach uses the trace of the covariance matrix instead of its determinant and, therefore, efficiently computing the overall gain. This is possible thanks to having the same representation units for all the observable features and, consequently, avoiding scalability problems. Finally, and despising the constants, the entropy of a discrete random variable can be efficiently computed as:

$$H(x) = \sum_{i=0}^{3n} \log(\Sigma_{ii}). \quad (3)$$

6.1 Scene Representation: 3D Occupancy Grid

A 3D occupancy grid is a map of a 3D space represented by a set of random variables, which are uniformly distributed on a discrete grid. These random variables are binary and specify whether each of the grid cells is occupied or free. Usually occupancy grid maps are used for building a consistent map after solving the SLAM problem, since

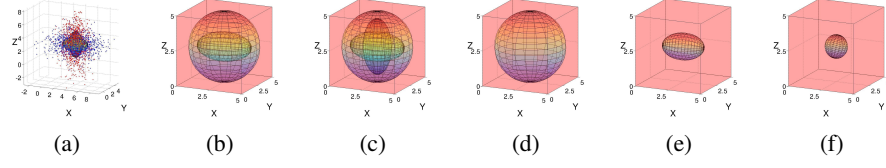


Fig. 3: Graphical interpretation through ellipsoids of the covariance matrix reduction inside a voxel. Figure a) shows two independent simulated readings of a point in space, which are taken to be perpendicular for clarity. Figure d) shows the a priori uncertainty of an unknown voxel represented as a covariance matrix and visualized as a sphere inscribed inside the voxel cube. Pairs of figures (b-e) and (c-f) show how the covariance matrix of a voxel gets updated after combining one or both readings, respectively.

they assume exact robot’s pose information. In a different way, our approach does not use the occupancy grid map as a final result but as a tool to evaluate the information gain of multiple possible view poses.

Our 3D occupancy grid map is based on a probabilistic voxel space defined by a multiresolution octree structure. All 3D grid cells, also called voxels, have associated a covariance matrix depending on all the history of measurements. At the same time, each voxel is defined by three possible occupancy types: *occupied*, *free* or *unknown*. By using the covariance matrix as an uncertainty voxel-related measurement, our approach can optimally obtain the information gain taking into account the orientation of the sensor. This is an important feature when using a noisy sensor such as a 3D ToF camera, since the error is usually bigger on one component.

6.2 Expected Gain Using an Occupancy Grid

Initially, all voxel states are set to *unknown*, state with the highest uncertainty. Once new sensor data are obtained, the states of all voxels intersected by a ray are updated. Depending on whether a voxel is crossed by a ray-trace or whether it encloses a new measurement, the voxel state is set to *free* or to *occupied*, respectively. Also, each *occupied* voxel is assigned its measurement covariance matrix Σ_i in order to posteriorly compute the information gain of new viewpoints. If the voxel was previously defined *occupied*, both the new covariance matrix and the former are combined as shown in Fig. 3 by

$$(\Sigma_i)^{-1} = (\Sigma_i^{t-1})^{-1} + (\Sigma_i^t)^{-1}. \quad (4)$$

Only voxels with *unknown* and *occupied* states would be considered for estimating the information gain, since *free* voxels do not provide any information. The reason for this behaviour is to minimize the effect of non-filtered noise and possible miss-readings due to non-systematic 3D ToF camera errors. Once the viewpoint estimator recommends a set of n viewpoints, their expected information gain (IG) is computed. Every viewpoint is simulated by ray-tracing from the sensor’s pose to the occupancy

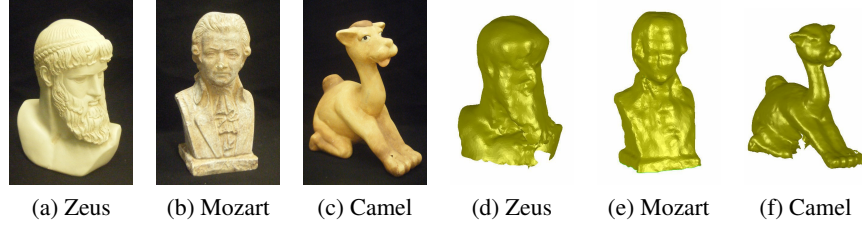


Fig. 4: a), b) and c) are the original object figures. d), e) and f), final triangle mesh of the modeled objects. Note that some details of the objects can not be captured due to the low resolution of the 3D ToF camera.

grid. Each colliding ray updates the corresponding voxel's covariance matrix and a copy is kept in memory as a sparse matrix

$$\mathbf{A} = \begin{pmatrix} \Sigma_0 & 0 & \cdots & 0 \\ 0 & \Sigma_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Sigma_n \end{pmatrix}. \quad (5)$$

Finally, the overall expected information gain is computed as

$$\mathbf{IG} = \sum_{i=0}^{3n} \log(\mathbf{A}_{ii}), \quad (6)$$

7 Experimentation

In order to be evaluated, the proposed active view planning has been applied to the task of modelling three free-form objects (see Fig. 4). Although all objects have similar sizes, each of them has its own degree of complexity, mainly based on the number of concavities. The most complex object is the Camel, followed by the Zeus and the Mozart bust. On the Mozart bust the only influential concavity is the one at its neck. The Zeus bust shows a higher complexity by having a big concavity under its beard. Finally, the Camel has a big concavity under its legs and a very heterogeneous structure due to its neck and head.

7.1 Setup

The current approach has been tested using a 7 DoF manipulator robot type Kuka KR16 and a 3D ToF camera type Swissranger SR4000 attached to its flange. The 3D ToF camera is attached in a 90 degree angle with respect to the tool-center-point (TCP). During the experiments, the objects were placed on a fixed and static platform at a height of approximately 670 mm. At this height, and due to its wide workspace, the Kuka KR16

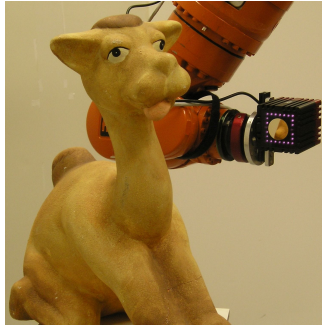


Fig. 5: Experimental setup. A SR4000 3D ToF camera is attached to the end effector of a Kuka KR16 industrial robot.

manipulator is able to cover comfortably the surrounding volume of a medium-sized object at a distance of 40 cm (see Fig. 5). The high accuracy of Kuka KR16 is required since the approach only takes into account the uncertainty of the measurements and not the one of the sensor's pose. Alternatively, if a high accuracy arm it is not available consecutive point clouds can be put in correspondence with a minimisation proces, as is done a.e. in [4]. The 3D ToF camera is intrinsically and extrinsically calibrated. Moreover, depth calibration is applied to improve 3D ToF camera measurements following the methodology of [7]. All together helps to get more accurate point clouds and to correctly register them in order to make the model grow.

7.2 Results

The Zeus, Mozart and Camel objects are appropriate for the evaluation because they have been used previously in other works. Moreover, the obtained results can be compared with two previous approaches used at DLR in the past (Fig. 6). Although a straightforward comparison between them is not an easy task because each approach used different sensors, some interesting conclusions can be extracted.

First, [8] used two different 3D ToF cameras, a Swissranger SR-3000 and IFM O3D100, for modelling the Camel. Their approach consisted in building a surface mesh by registering a pre-defined human-driven set of measurements and merging the views using Iterative-Closest-Point (ICP) without possibly refine their result. Their resulting models are reproduced in Fig. 6a and 6b. Later, Kriegel [10] used a laser scanner for modelling both the Camel and Mozart objects, providing a very precise model at the expense of a time consuming process. Fig. 6d shows the final model of the Camel.

Alternatively, our proposed algorithm (Fig. 6c) presents a good model compromise between model precision and time acquisition for most robotic tasks. Our approach takes explicit advantage of the noisy and low resolution data obtained from a 3D ToF camera. Images are obtained at a high frame rate (20 fps.) and with less complexity in the setup. Note that with this algorithm an explicit uncertainty measure of each point of the model is maintained, and therefore it is easy to define measures of the overall quality on the model. One natural consequence of the algorithm is that for complex

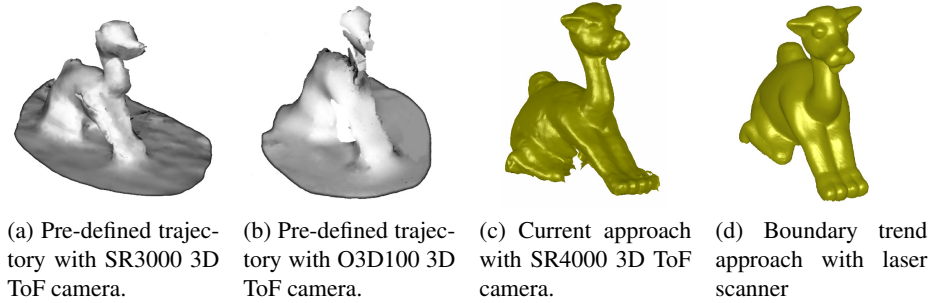


Fig. 6: a) and b) from [8]. c) is our current approach and d) from [10].

objects with concavities and details, a higher number of view points for completing the model is required and automatically computed. On the contrary, the simpler the object, the fewer number of views. But more important, the algorithm can intrinsically refine the model to a desired precision, always limited by the sensor’s resolution, by defining a threshold on the overall amount of information gain.

Figure 7 shows a graphical example of how candidate views are computed and how the model is incrementally updated. Step by step, the algorithm adds new measurements to the model based on the maximum information gain. Those new measurements are previously forced to belong to a contiguous area. By applying these two constraints, the algorithm succeeds to model almost completely any free-form objects. It can be seen that the Zeus bust and the Camel have a hole on their surfaces. These holes are a consequence of the sensor’s configuration and the impossibility of the robot arm to attain the required poses. Figure 4 shows the final result of each of the objects used in three of our experiments.

8 Conclusions and Future Research

The paper presents a hybrid active view planning approach and its application to autonomously modelling an unknown free-form 3D object using a noisy sensor. The method combines viewpoint generation and viewpoint selection based on evaluation of the information gain. The method has been evaluated experimentally using a calibrated 3D ToF camera and a robotic arm on the task of 3D object modelling.

The proposed algorithm keeps track explicitly of the uncertainty in each point of the space. Using the proposed information gain measure as a criterion for evaluating the different views provides a trade-off between vicinity exploration and model refinement. Moreover, by keeping the complete covariance matrix in every voxel of the 3D occupancy grid, our method allows not only to reduce the uncertainty over the already seen voxels but to compute the information gain taking also into account the orientation of the sensor’s pose. This naturally encodes the idea that to obtain better information the same point has to be observed from different points of view, as has been shown. Observe that it is very important to calibrate the sensor and characterize its inherent

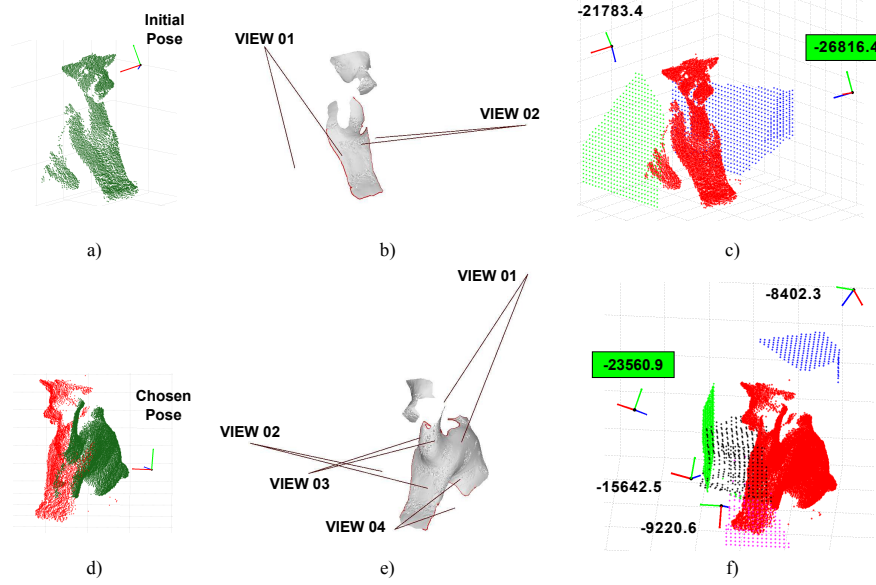


Fig. 7: Graphical representation of the steps carried out to compute consecutive view-points. a) shows the initial acquired point cloud. b) shows the corresponding mesh and the two candidate viewpoints on the basis of the detected edge-trends. c) simulates the measurements from the previous viewpoint candidates and the one with maximum information gain is chosen (marked on green). d) shows how the new point cloud is integrated into the previous one. e) shows the new corresponding mesh and the four new candidates. Finally, f) shows the new simulated ray-tracing measurements. Observed how part of the simulated ray-tracings on VIEW 03 do not impact on the not-sensed bounding box but into areas already seen.

uncertainty, as this is a lower bound measure of the uncertainty in each point of the model.

The algorithm keeps track of the overall uncertainty of the model, and it is possible also to envisage ways to compute the overall uncertainty of selected parts. In the future, this algorithm can be used to build multiresolution models by roughly modelling some parts and precisely modelling other parts. Obviously this would be useful for applications, such as grasping, insertion or part removal.

9 Acknowledgments

This work has been partially supported by the EU IntellAct project FP7-ICT2009-6-269959, the Spanish Ministry of Science and Innovation under the project DPI2011-27510, and by KUKA Roboter GmbH. S. Foix is supported by PhD fellowship from CSIC's JAE program. The authors gratefully acknowledge the suggestions and deep insight of Michael Suppa and Tim Bodenmüller.

References

1. Alenyà, G., Dellen, B., Torras, C.: 3D modelling of leaves from color and tof data for robotized plant measuring. In: Proc. IEEE Int. Conf. Robot. Automat. pp. 3408–3414. Shanghai (May 2011)
2. Chen, S., Li, Y., Zhang, J., Wang, W.: Active sensor planning for multiview vision tasks. Springer-Verlag, Berlin Heidelberg (2008)
3. Deinzer, F., Derichs, C., Niemann, H.: A framework for actively selecting viewpoints in object recognition. *Int. J. Pattern Recogn. Artif. Intell.* 23(4), 765–799 (2009)
4. Foix, S., Alenyà, G., Andrade-Cetto, J., Torras, C.: Object modeling using a ToF camera under an uncertainty reduction approach. In: Proc. IEEE Int. Conf. Robot. Automat. pp. 1306–1312. Anchorage (May 2010)
5. Foix, S., Alenyà, G., Torras, C.: Lock-in Time-of-Flight (ToF) cameras: a survey. *IEEE Sensors J.* 11(9), 1917–1926 (Sep 2011)
6. Foix, S., Alenyà, G., Torras, C.: Towards plant monitoring through next best view. In: Proc. 14th Int. Conf. Cat. Assoc. Artificial Intelligence. Lleida (Oct 2011)
7. Fuchs, S., Hirzinger, G.: Extrinsic and depth calibration of ToF-cameras. In: Proc. 22nd IEEE Conf. Comput. Vision Pattern Recog. vol. 1-12, pp. 3777–3782. Anchorage (June 2008)
8. Fuchs, S., May, S.: Calibration and registration for precise surface reconstruction with time of flight cameras. *Int. J. Int. Syst. Tech. App.* 5(3-4), 274–284 (2008)
9. Jia, Z., Chang, Y., Chen, T.: A general boosting-based framework for active object recognition. In: Proc. British Machine Vision Conf. pp. 46.1–46.11. BMVA Press (Sep 2010)
10. Kriegel, S., Bodenmüller, T., Suppa, M., Hirzinger, G.: A surface-based next-best-view approach for automated 3D model completion of unknown objects. In: Proc. IEEE Int. Conf. Robot. Automat. pp. 4869–4874. Shanghai (May 2011)
11. Madsen, C., Christensen, H.: A viewpoint planning strategy for determining true angles on polyhedral objects by camera alignment. *IEEE Trans. Pattern Anal. Machine Intell.* 19(2), 158–163 (Feb 1997)
12. Maver, J., Bajcsy, R.: Occlusions as a guide for planning the next view. *IEEE Trans. Pattern Anal. Machine Intell.* 15(5), 417–433 (May 1993)
13. May, S., Droschel, D., Holz, D., Fuchs, S., Malis, E., Nuchter, A., Hertzberg, J.: Three-dimensional mapping with time-of-flight cameras. *J. Field Robotics* 26(11-12), 934–965 (Nov-Dec 2009)
14. Ramisa, A., Alenyà, G., Moreno-Noguer, F., Torras, C.: Determining where to grasp cloth using depth information. In: Proc. 14th Int. Conf. Cat. Assoc. Artificial Intelligence. Lleida (Oct 2011)
15. Roy, S.D., Chaudhury, S., Banerjee, S.: Active recognition through next view planning: A survey. *Pattern Recogn.* 37(3), 429–446 (2004)
16. Saxena, A., Sun, M., Ng, A.Y.: Make3d: Depth perception from a single still image. In: Proc. 23th AAAI Conf. on Artificial Intelligence. pp. 1571–1576. Chicago (Jul 2008)
17. Schiele, B., Crowley, J.: Transinformation for active object recognition. In: Proc. IEEE Int. Conf. Comput. Vision. pp. 249–254. Bombay (Jan 1998)
18. Scott, W.R., Roth, G.: View planning for automated three-dimensional object reconstruction and inspection. *ACM Computing Surveys* 35(1), 64–96 (Mar 2003)
19. Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423 (1948)
20. Sim, R.: Stable exploration for bearings-only SLAM. In: Proc. IEEE Int. Conf. Robot. Automat. pp. 2422–2427. Barcelona (Apr 2005)