Towards plant monitoring through Next Best View

Sergi FOIX [1], Guillem ALENYÀ and Carme TORRAS
Institut de Robòtica i Informàtica Industrial, CSIC-UPC

Abstract. Monitoring plants using leaf feature detection is a challenging perception task because different leaves, even from the same plant, may have very different shapes, sizes and deformations. In addition, leaves may be occluded by other leaves making it hard to determine some of their characteristics. In this paper we use a Time-of-Flight (ToF) camera mounted on a robot arm to acquire the depth information needed for plant leaf detection. Under a Next Best View (NBV) paradigm, we propose a criterion to compute a new camera position that offers a better view of a target leaf. The proposed criterion exploits some typical errors of the ToF camera, which are common to other 3D sensing devices as well. This approach is also useful when more than one leaf is segmented as the same region, since moving the camera following the same NBV criterion helps to disambiguate this situation.

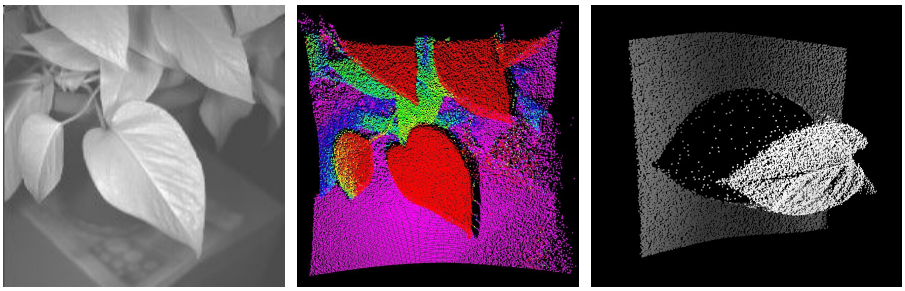
Keywords. Next Best View, ToF cameras, depth images, plant segmentation, leaves disambiguation

Introduction

Food industry is very important for society, and large areas of the world are currently cultivated, as open plantations or as greenhouses. The automation in such areas has been traditionally intensive, generally at large scale and relying on human assistance. Recently, more attention is given to standalone processes taking increasingly into account plants as individuals [1]. In the context of the GARNICS project, we aim at the monitorization of large plantations to help determine the best treatments (watering, nutrients, sunlight) to optimize pre-defined aspects (growth, seedling, flowers) and eventually guiding robots to interact with plants in order to obtain samples from leaves to be analysed or even to perform some prunning.

Monitoring and taking actions over plants are two very difficult tasks. The reason why these tasks are so difficult is because plants are complex and dynamic systems. Two plants are not equal. They are composed of multiple elements such as flowers, leaves, stem and roots. They grow, changing their shape and incorporating new elements. They move and they change their colors depending not only on intrinsic but also extrinsic components. Because of all of these plant behaviours, tasks such as feature detection and action planning over them are very hard problems to solve.

[1]Corresponding Author: Sergi Foix, Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain; E-mail: sfoix@iri.upc.edu



(a) Intensity image (b) Depth colored 3D point cloud (c) False depth points at edges

Figure 1. Typical images acquired with a ToF camera (200×200 PMD CamCube 3.0). Interesting false depth measures appear at the edges between foreground and background due to the integration of the reflected light of both surfaces in the corresponding pixels.

But leaves are not uniformly arranged in space, albeit they grow, by their nature, in a very structured way. Therefore, monitoring and measuring actual properties of the plant over its leaves requires of specific tecniques in order to place a sensor into the correct pose. The next-best-view algorithm presented in this article focuses its attention to find a point of view that provides a better perception of an occluded leaf or, in a similar way, to disambiguate about the quantity of observed leaves. At the same time and as a consequence, a better estimation of the leaves poses is achieved, a necessary requirement to achieve the placement of a tactile measuring tool over the leaf.

Plants are a hard scenario for segmentation algorithms based on traditional color vision, mainly due to the lack of texture and the uniformity of color. It has recently been demostrated that 3D information is highly valuable in this context [2]. Such information is obtained with a depth sensor, that should provide information independently of the illumination conditions, as they change in greenhouses. Acquisition time is also important, as a lot of plants should be monitorized. Finally, the sensor has to be lightweight, as we want to mount it in the end effector of a robotized arm. Time-of-Flight cameras are lightweight 3D cameras that provide directly depth images without pre-processes, with infrared autoillumination units, that deliver 30 frames per second.

The article is structured as follows: in Sec. 1 the 3D image acquisition through ToF cameras is introduced. Section 2 explains the proposed Next-Best-View algorithm and how it takes advantage of erroneously captured data. This algorithm is validated in Sec. 3, including also some considerations about the camera and scene configuration. Finally, in Sec. 4 conclusions and future work are presented.

1. 3D image acquisition

Depth measurements are carried out by a relatively new type of sensor named Time-of-Flight (ToF) camera. This type of sensor has the main characteristic of providing registered depth and intensity images of a scene at a high frame-rate (see Fig. 1(a) and 1(b)). ToF cameras use the well-known time-of-flight principle to compute depth. The camera emits modulated infra-red light in order to measure the travelling time between the known emitted waves and the ones reflected back over the objects in the scene. Compared to other similar technologies, such as the new Kinect, and taking into account the

context of the GARNICS project, ToF cameras provide some interesting features that make them more suitable for short range applications. It has auto-illumination, making it independent from external light sources, and its minimal depth measuring range can get as close as 15 cm. [2]

But ToF-cameras have two main drawbacks: low resolution (200×200 pixels for a PMD CamCube 3.0 camera) and noisy depth measurements due to systematic and non-systematic errors. On the one hand, low resolution can be a big problem for large environment applications, but it has not such a negative impact when the camera is used at 20 cm range as it is our case[3]. On the other hand, noisy depth measurements due to non-systematic errors get amplified by working in such a short range. Mainly the ones due to multiple light reception and light scattering. Systematic errors get highly reduced by calibration procedures [3]. For a more detailed and wide classification and explanation of the different error sources, advantages and limitations of ToF cameras, please refer to [4].

There is one type of multiple light reception error that deserves special attention in this article. This is the jump-edge error (Fig. 1(c)). This type of error appears due to the mix of measurements over the pixels that contain the edges between foreground objects and their background, refer to Sec. 2.2 for a more detailed explanation. Our approach takes advantage of detecting this type of error on the scene, and computes a new next-best-view in order to acquire a better estimation of the leaves composition. Jump-edge errors are not unique of ToF cameras but are also present in lidar systems and the new Kinect[4].

2. Improving 3D information through Next Best View

Next-best-view (NBV) is one of the most challenging problems in vision sensor planning. Its application covers tasks such as autonomous 3D object modelling, object recognition, visual tracking or, as in our case, monitoring complex systems. Initial investigations in the field of NBV were presented in [5], giving two algorithms to determine best next views that established the basis for further research: the planetarium algorithm (slower due to consider possible occlusions), and the normal algorithm (much faster, but weaker with occlusions). Subsequent research studied the use of camera triangulation systems, and in [6] the use of a range scanner was suggested. The authors concluded that using depth information from range data into the NBV problem was a tool for cost-effective and accurate acquisition of 3D data. More recently, in [7], a method is proposed for automatically acquiring 3D models of unknown objects by moving the sensor around the target object. Sensor motion is determined by the analysis of the curvature's trend at the surface edges.

The level of difficulty in NBV does not depend only on the task but also on some common aspects such as: whether a prior model of the object is known or not, whether a very precise range sensor is used or not, and whether the viewpoint working space is highly constrained or not. In this work we assume that plants are composed of nearly

[2] Measures extracted with a PMD CamCube 3.0 camera after changing its modulation frequency to 21MHz and decreasing its integration time to 0.2 ms.

[3] 20 cm ensures a good compromise between planar model fitting and signal-to-noise ratio.

[4] Due to its internal filtering, Kinect does not deliver these data.

planar leaves so we rely on planar models, a noisy 3D range sensor is used and the viewpoint working space is constrained by the manipulator robot working space and by the pre-defined maximum distance between the camera and the surface of the plant.

Although the following sections give a more comprehensive explanation of each of the steps in the view sensor planning, here is a brief summary. Initially, the camera is placed at approximately 15-20 cm away from the plant's region of interest. Secondly, leaves are segmented by means of planar approximation. Thirdly, jump-edge points are detected. And finally, by combining the data from the previous two steps, the NBV is computed.

2.1. Leaf segmentation - Fitting planar models to leaves

Each plant has its own specific type of leaves and their shapes and sizes can vary in a wide range. Although more accurate leaf 3D models can be defined and consequently improve the detection of leaves and the estimation of their poses, in our approach a simple planar model has been used. Fitting accurate 3D object models to crowded scenes is a very time consuming task, and it gets worse when the data provided are noisy as it happens in the case of ToF cameras. Consequently, and when plants have nearly planar leaves, simple plane models can be approximated and therefore increase the speed of 3D data processing.

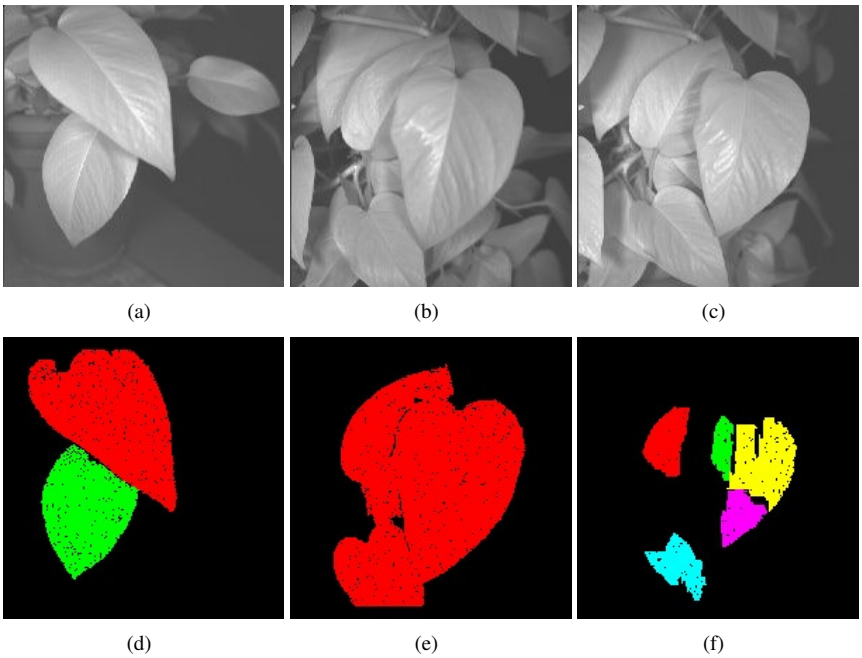
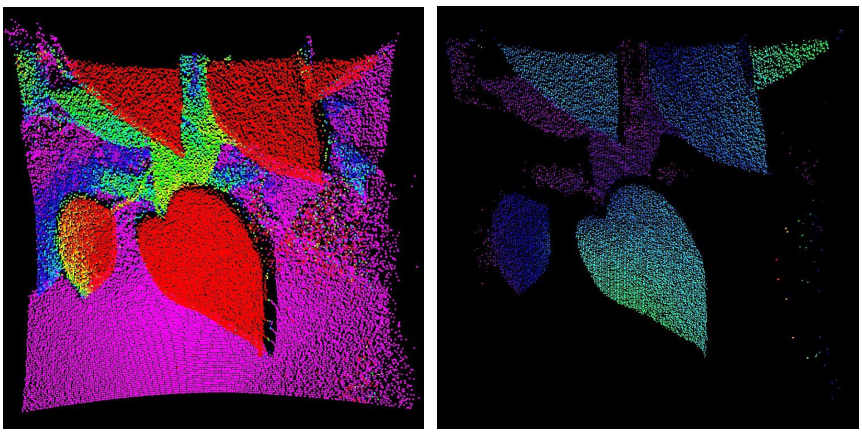


Figure 2. Planar leaf segmentation is highly parameter-dependent. The first row shows some intensity images, while the second row shows their corresponding planar segments defined by colors. Images (a,d) and (b,e) share the same parameterization. It is possible to see how we obtain different segmentation results for the same parameterization. Images (c,f) are the same scene as (b,e) but with different parameterizations. Here it is possible to see how a bad leaf segmentation is produced due to the non-planar shape of one of the leaves.

But there are always some drawbacks. Planar leaf segmentation is a highly parameter-dependent algorithm. Depending on the shape of the sensored surfaces that need to be modelled and the quality and density of the acquired 3D data, necessary pre-processes for plane estimation, such as point-normal calculation and point-neighbourhood computation, can be very tricky to tune. In the case of plants with planar leaves, where data is captured with a ToF camera, these tunning parameters have to allow dealing with the highly noisy readings from the sensor and try not to subdivide a single leave in multiple planes. An example of a bad parameterization can be observed in Fig. 2(f). It is preferable fusing two leaves as if they were a single one than subdividing a single leaf in sub-elements. This is because, as it has been said previously and will be demonstrated by experiments in Sec. 3, ambiguity can be resolved by acquiring a new best view.

2.2. Jump-edge filter

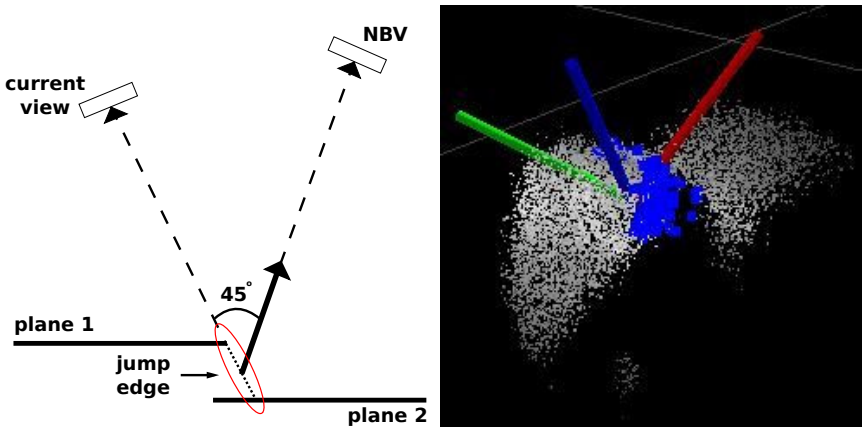
Figure 3 shows the appearance of a curtain of flying points around the edges between foreground objects and their background. These points are commonly known as jump-edge points and are generally removed by comparing the angle of incidence of neighboring pixels [8,9,10]. They are false measurements and consequently they are always removed from the data sets, even the new Kinect sensor filters internally these misreadings. But in our case the appearance of these false measurements are indicative of possible model misinterpretation or object occlusion. Therefore, their detection and 3D localization in the scene provide the required information for computing the next-best-view that will try to disambiguate or improve occluded leaf visibility and pose estimation. In our algorithm, a number of at least 20 jump-edge points have to be detected in order to consider them a region of interest. This threshold has been set empirically to prevent considering non-systematic noise as jump-edge points.



(a) Raw 3D points colored by depth (b) Filtered 3D points colored by depth

Figure 3. Comparison between raw and filtered 3D point clouds. Image (a) clearly shows how raw measurements incorporate undesired data into the 3D point cloud. A curtain of points can be identified on the edges between the foreground (leaves) and the background. Image (b) shows the 3D point cloud after the jump-edge and bounding-box filters have been applied.

2.3. Next position computation



(a) 2D NBV schematic representation (b) Jump-edge points (Blue)

Figure 4. False depth measurements (jump-edge points) detection helps to compute the NBV to uncover occluded leaves and to disambiguate the number of observed leaves. Figure (a) shows the 2D schematic representation of the algorithm. Figure (b) shows, in blue, the 3D jump-edge points.

Figure 4(a) shows a schematic representation of the computation of the NBV for the tasks of uncovering occluded leaves and leaves disambiguation. The main characteristic of our NBV method is that it takes advantage of erroneous depth readings (Fig. 4(b)) for computing a better view in a geometrical way.

Once the overall estimated planes and jump-edge points have been obtain, the computation of the NBV is reduced to a geometrical problem. As introduced in previous sections, the NBV is only calculated if there are jump-edge pixels adjacent to two planes or if these are contained inside a unique plane. For any of both conditions the algorithm behaves in the same manner. First, the median point of the jump-edge points that fulfill the condition is calculated and normalized as a unitary vector. This vector represents the current view camera direction. Second, we calculate the cross product between the estimated plane normal [5] and the previous normalized vector. The resulting orthonormal vector is the one that will act as a rotation axis to attain the NBV (on the schematic representation, this vector would come out from the figure). Finally, using the median jump edge point as a center and the previous rotation axis, a rotation of 45 degrees is applied to the current view. Although 45 degrees have proven to be an adequate quantity in our experiments, it is advisable to use smaller angles, e.g. 10 degrees, since incremental NBV is more adaptative. It has to be noticed that the current method guarantees a gain of information over the scene on superficial leaves but not on the ones deep inside the plant, since their probability of being occluded by unobserved leaves is very high.

[5] In the task of resolving leaf occlusion the normal vector is the one of the occluding plane (closer to the camera).

Figure 5. WAM arm used in the experiments holding the Time-of-Flight camera observing a plant.

3. Experiments

Figure 5 shows the experimental setup of our simulated monitoring plant process. It includes a PMD CamCube 3.0 ToF camera mounted as an end-effector of a 7-DoF Barrett WAM arm. This configuration permits moving the camera to different viewpoints and also monitoring several plants located in the typical matrix-like plant containers.

As it has been previously stated, our proposed NBV algorithm has been designed in order to deal with two specific tasks, resolution of leaves occlusions and disambiguation between leaves. Figures 6 and 7 show two scenes where both tasks have been performed respectively. Each figure is divided in two sets of images, the images at the top row show the state of the scene before applying the NBV algorithm while the images at the bottom row show its state afterwards. By observing the intensity images of the plant it is easy to imagine how common these two types of scenes are obtained in a plant monitoring process and, consequently, how important it is to be able to deal efficiently with occlusions and ambiguities.

Figure 6(a) shows the intensity image of a scene where the occlusion of a leaf is clearly identified. By executing the jump-edge filter over the 3D data, the countours of each leaf are extracted (Fig. 6(b)). At the same time, the plane segmentation process provides the estimation of the different planes (Fig. 6(c)). Figure 6(d) shows, in a 3D rotated view, the extracted jump-edge points that fall just in the frontier between both leaves. These points are the ones that allow us to compute the NBV whose result is displayed at the bottom row of Fig. 6. By comparing the image pairs Fig. 6(a, e) and Fig. 6(c, g), it can be seen by moving the camera to the NBV the overall perception of the occluded leaf surface is significally improved.

Figure 7 shows the ambiguity scene where two leaves have been misinterpreted as only one. In order to evaluate whether there is an ambiguity, the existence of jump-edge points inside the segmented plane is verified. Fig. 7(b) shows how part of the jump-edge points, white contours, are found inside the area of the wrongly assumed leaf (Fig. 7(c)). Following the same NBV approach as before, a new camera pose is computed leading to the resulting images at the bottom row. After the robot's movement, the previously estimated dark red plane has now been correctly divided into two different planes, as it

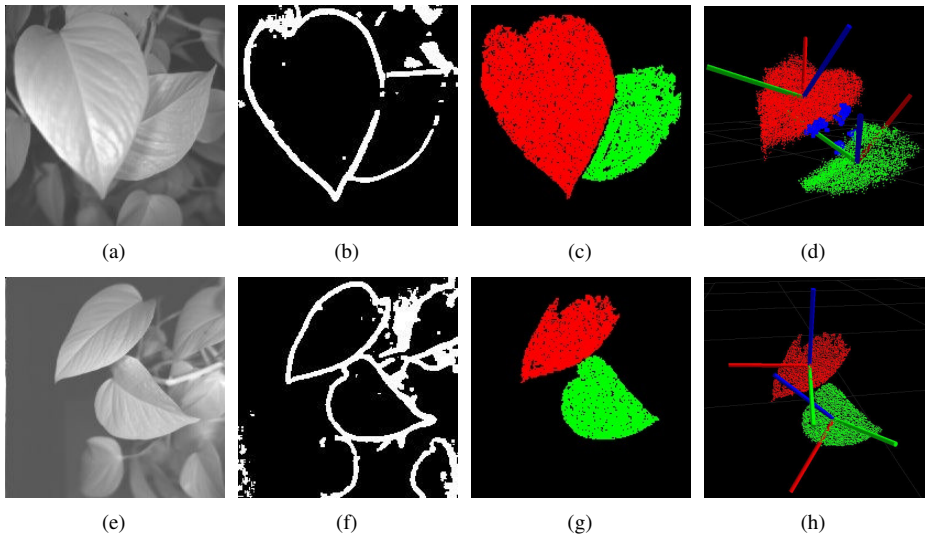


Figure 6. Scene containing a detected leaf occlusion. Top row shows the scene before applying the NBV algorithm, images (a-d). Bottom row shows the scene observed from the new viewpoint, images (e-h). After applying the NBV algorithm the occluded leaf is clearly discovered.

was expected (Fig. 7(g)). Figures 7(d, h) show the final 3D point cloud of the leaves as if they were viewed from the same camera pose, before and after the NBV. It can be clearly seen how not only the disambiguation has been achieved but also how part of one of the leaves that was occluded is now uncovered.

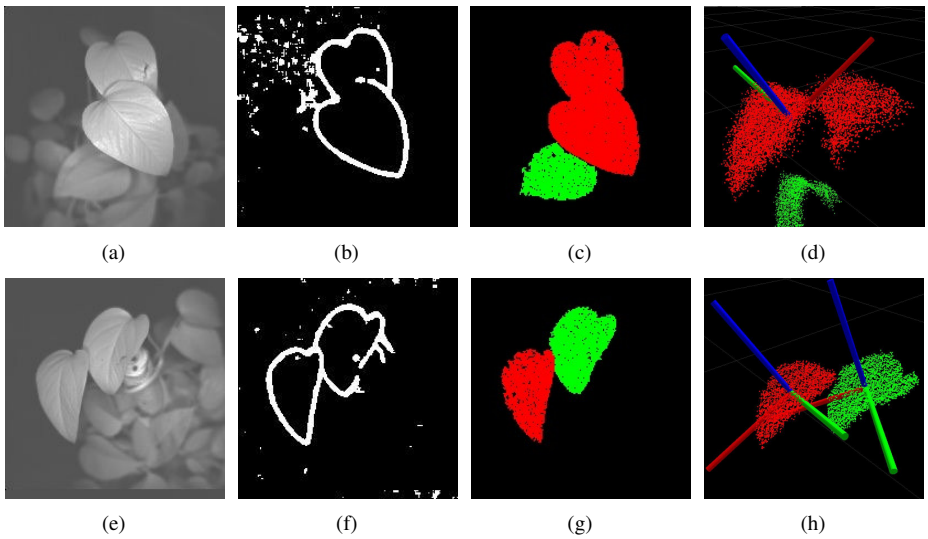


Figure 7. Scene containing a possible mixture of leaves. Top row shows the scene before applying the NBV algorithm, images (a-d). Bottom row shows the scene after it, images (e-h). After applying the NBV algorithm the ambiguity is clarified and two leaves are detected instead of one.

4. Conclusions and future work

This paper proposed a novel method to efficiently estimate a NBV for improving plant monitoring. The method takes advantage of jump-edge flying points, typical erroneous data from a ToF camera, for finding a suitable solution to two common monitoring tasks, getting a better view of an occluded target leaf and resolving ambiguity in the number of leaves. The method can be executed in real-time since it does not use any cost function minimization approach or any complex leaf model fitting but a geometrical approach and a simple planar leaf model.

It has to be noticed that, depending on the configuration of leaves, it may not be possible to completely avoid occlusions or ambiguities by moving the camera. Next research steps will focus on using robot manipulation to help monitoring tasks.

Acknowledgements

This research is partially funded by the EU GARNICS project FP7-247947, by the Spanish Ministry of Science and Innovation under projects DPI2008-06022 and MIPRCV Consolider Ingenio CSD2007-00018, and the Catalan Research Commission through the Robotics Group. S. Foix is supported by a PhD fellowship from CSIC's JAE program.

References

[1] R.D. King, J. Rowland, S.G. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L.N. Soldatova, A. Sparkes, K.E. Whelan, and A. Clare. The automation of science. *Science*, 5923(324):85–89, 2009.

[2] G. Alenyà, B. Dellen, and C. Torras. 3d modelling of leaves from color and tof data for robotized plant measuring. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 3408–3414, Shanghai, May 2011.

[3] S. Fuchs and G. Hirzinger. Extrinsic and depth calibration of ToF-cameras. In *Proc. 22nd IEEE Conf. Comput. Vision Pattern Recog.*, volume 1-12, pages 3777–3782, Anchorage, June 2008.

[4] S. Foix, G. Alenyà, and C. Torras. Lock-in Time-of-Flight (ToF) cameras: a survey. *IEEE Sensors J.*, 2011. to appear.

[5] C. I. Connolly. The determination of next best views. In *Proc. IEEE Int. Conf. Robot. Automat.*, volume 2, pages 432–435, St. Louis, Mar. 1985.

[6] Y. Zhien, W. Ke, and Y. Rong-Guang. Next best view of range sensor. In *IEEE 22nd International Conference on Industrial Electronics, Control, and Instrumentation (IECON)*, volume 1, pages 185–188, Aug. 1996.

[7] 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. Robot. Automat.*, Shanghai, May 2011.

[8] S. Fuchs and S. May. 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] T. Kahlmann and H. Ingensand. Calibration and development for increased accuracy of 3D range imaging cameras. *J. Appl. Geodesy*, 2(1):1–11, 2008.

[10] W. Karel, P. Dorninger, and N. Pfeifer. In situ determination of range camera quality parameters by segmentation. In *Proc. 8th Int. Conf. on Opt. 3D Meas. Tech.*, pages 109 – 116, Zurich, July 2007.