# ToF cameras for active vision in robotics

G. Alenyà[a,*], S. Foix[a], C. Torras[a]

[a]*Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain*

**Abstract**

ToF cameras are now a mature technology that is widely being adopted to provide sensory input to robotic applications. Depending on the nature of the objects to be perceived and the viewing distance, we distinguish two groups of applications: those requiring to capture the whole *scene* and those centered on an *object*. It will be demonstrated that it is in this last group of applications, in which the robot has to locate and possibly manipulate an object, where the distinctive characteristics of ToF cameras can be better exploited.

After presenting the physical sensor features and the calibration requirements of such cameras, we review some representative works highlighting for each one which of the distinctive ToF characteristics have been more essential. Even if at low resolution, the acquisition of 3D images at frame-rate is one of the most important features, as it enables quick background/foreground segmentation. A common use is in combination with classical color cameras. We present three developed applications, using a mobile robot and a robotic arm, to exemplify with real images some of the stated advantages.

*Keywords:* Time-of-Flight cameras; 3D perception for manipulation; depth calibration; outdoor imaging; complex-shape objects

## 1. Introduction

A Time-of-Flight (ToF) camera is a relatively new type of sensor that delivers 3-dimensional images at high frame rate, simultaneously providing intensity data and range information for every pixel. In robotics it has been used for a wide range of applications that can be classified in two groups: scene-related and object-related. In scene-related tasks usually the camera is mounted on a mobile robot (Fig 1a) and it is used for mapping and localization, although it has been also applied to obstacle detection and terrain classification [1]. In object-related tasks, the camera is usually attached to the end-effector of a robot manipulator (eye-in-hand configuration), so that new images can be obtained by actively changing the point of view of the camera (Fig. 1b). In this paper we will describe some of the lessons learned in using such cameras to perform robotic tasks, with special attention to eye-in-hand configurations.

In an eye-in-hand scenario, some particular characteristics of the sensor system are appreciated. Mainly, the compactness and the detection in a short range, besides the obvious requirement of quality (precision and accuracy) in the obtained data. On the one hand, operation in a short range is desired because robot manipulators have typically a limited workspace, and the distance from the end-effector to an object located in front of the robot is short. As will be demonstrated later, ToF cameras exhibit good performance in short ranges. On the other hand, as the sensor system is mounted on a robot arm it has to be lightweight, with no mobile parts, and as small as possible to avoid interference with the environment or the robot itself. ToF cameras fit well this description, as they are usually lightweight, have no mobile parts, and they can be compact and small as well. Section 2 introduces ToF cameras and presents a critical comparison with RGBD cameras (Kinect), a different 3D sensor that is more and more commonly used in robotics.

Section 3 presents and puts in context some of the relevant works, distinguishing between scene-related tasks, where the camera is used for scene understanding, and object-related tasks, which involve the detection, the pose estimation, or the manipulation of objects. Our aim is to highlight in each case the main characteristics of ToF cameras that are differential compared to other approaches.

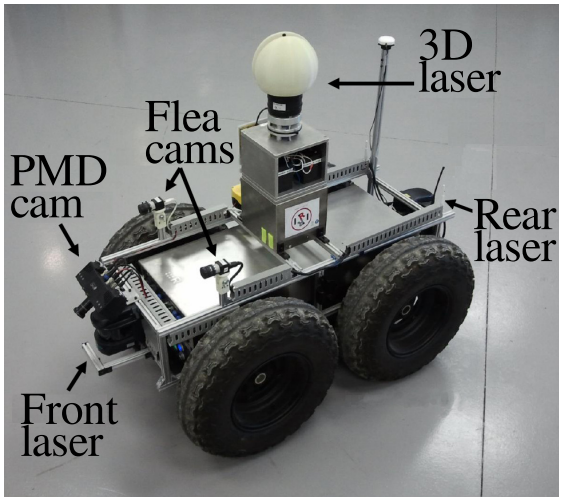
Regarding the quality of data, it is well known that raw ToF data is quite noisy and prone to several types of disturbances [2]. Some of them are systematic and can be calibrated, and others are non-systematic and sometimes can be filtered out. In Section 4 systematic and non-systematic error sources are reviewed and a calibration algorithm is proposed that takes into account the application scenario.

The ability to actively move the camera depending on the scene provides some advantages. In Section 5 we show three illustrative examples: understanding the 3D structure of some relevant parts of the scene to enable robot-object interaction with plants, obtaining detailed views of relevant parts of textiles objects, and disambiguation to enhance segmentation algorithms. Finally, some conclusions are drawn in Section 6.

## 2. ToF cameras

In ToF cameras depth measurements are based on the well known time-of-flight principle. A radio frequency modulated light field is emitted and then reflected back to the sensor,

*Corresponding author
*Email address:* galenya@iri.upc.edu (G. Alenyà)




(a) Scene-related task setup: ToF camera attached at the front of a mobile robot used for navigation and terrain classification. Reproduced from [1]

(b) Object-related task setup: ToF camera attached to a manipulator end-effector, in this case a Chlorophyll meter, to get measures by touching the leaf surface.

Figure 1: Camera-robot configurations for active vision.

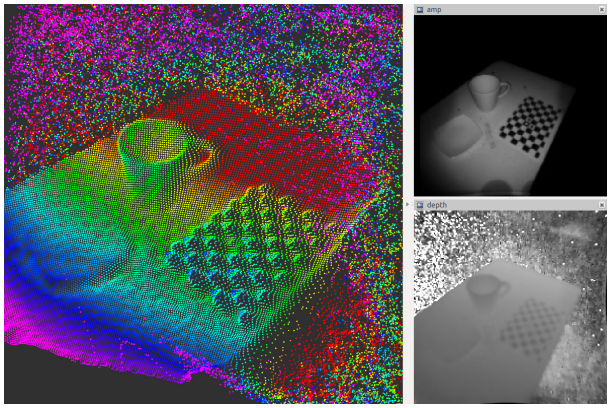


Figure 2: Typical raw ToF image of a table with some objects at short distance. Right and above: intensity image. Right and down: Depth codified as intensity. Left: rotated 3D point cloud with depth color coded. Observe the errors in depth due to the colors in the calibration pattern and the noise in the background.

which allows for the parallel measurement of its phase (cross-correlation), offset, and amplitude [3]. Figure 2 shows a typical raw image of a flat surface with the depth values coded as color values.

The main characteristics of two ToF sensors, PMD CamCube 3 and Mesa Swissranger 4K are detailed in Table 1. We include also the specifications of the Kinect sensor to compare with a very common alternative 3D sensor. Both camera types can deliver depth images at reasonably high frame rates. The main difference is in resolution: ToF cameras still have limited resolution (typically around 200 x 200), while the Kinect depth camera exhibits VGA resolution. Both camera types are auto-illuminated so in principle they can work in a wide variety of illumination conditions.

In this paper we focus on 3D perception for robotic manipulation and object modeling, thus resolution is an important factor. It is worth mentioning that the closest working depth for Kinect is 0.5m[1] whereas that for ToF can reach 0.3m, and even 0.2m when equipped with new illumination units [2]. Kinect resolution is higher but closer views can be obtained with ToF cameras. Consequently, the resulting horizontal (or vertical) resolution in mm per pixel of both cameras is very similar as the lower resolution of ToF cameras can be compensated with closer image acquisition. The major consequence is that the density of the point cloud when viewing a given object is similar for both camera types.

However, placing ToF cameras closer to the object has two problems, related to focus and integration time, respectively. Like any other camera that uses optics, focus determines the depth of field (distance range where sharp images are obtained). If we set the focus to obtain sharp images of closer objects then the depth of field is small. ToF cameras do not have auto-focus

[1] It is commonly accepted that 0.7m is the closest distance, but in our tests we have been able to obtain depth images at 0.5m. New Kinect camera, to appear in the beginning of 2014 is supposed to work at 0.3m.

[2] http://www.pmdtec.com, 2013

| Camera model | PMD CamCube | Swissranger 4K | Kinect |
|---|---|---|---|
| Technology | ToF | ToF | Structured light |
| Image size | 200x200 | 176x144 | 640x480 (depth)<br>1280x1024 (color) |
| Frame rate | 40 fps<br>up to 80fps | 30 fps<br>up to 50fps | 30fps (depth)<br>30/15fps (color) |
| Lens | CS mount f = 12,8 | Standard/Wide option | fixed |
| Range | 0.3 - 7m | 0.8 - 5m<br>0.8 - 8m | 0.5 - 3.5m |
| Field of view | 40x40 | 43.6x34.6<br>69x56 | 57x43 |
| Focus | Adjustable | Adjustable | Fixed |
| Integration time | Manual | Manual | Auto |
| Illumination | Auto | Auto | Auto (depth) |
| Outdoor | Suppression<br>Background<br>Illumination | No | No |
| Images | Depth<br>Intensity<br>Amplitude<br>Confidence | Depth<br>Intensity<br>Amplitude<br>Confidence | Depth<br>Color |
| Interface | USB | USB - Ethernet | USB |

Table 1: Specifications of two different ToF cameras, and comparison with Kinect features.

capabilities, so the focus (and consequently the desired depth of field) has to be determined in advance.

Moreover, integration time has to be manually adjusted. Integration time has a strong impact on the quality of the obtained images, and each integration time sets the camera for a specific range of depths. As before, for close distances the range of possible depths for a given integration time is small.

Some of the ToF cameras have the capability of auto-adjusting the integration time. However, depth calibration of ToF cameras is dependent on the current integration time, and a common practice is to calibrate for only one integration time, which is manually determined depending on the expected depth range.

One of the advantages of Kinect is the ability of delivering colored depth points if required. Coloring ToF depth points is also possible but requires some additional efforts.

One common problem with both cameras is that they do not provide a dense depth map. The delivered depth images contain holes corresponding to the zones where the sensors have problems, whether due to the material of the objects (reflection, transparency, light absorption) or their position (out of range, with occlusions). Kinect is more sensitive to this problem by construction.

Finally, we have tested ToF cameras in outdoor scenarios with sunlight [4]. An algorithm has been proposed to select the best integration time depending on the sun conditions, as well as a suitable strategy to combine two frames to obtain depth images even when a plant is partially illuminated with direct sunlight and partially in shadow, as it is common in greenhouses. As could be expected, a ToF camera provides depth information but with more noisy depth readings in parts exposed to direct sunlight.

## 3. Using ToF cameras in robotic manipulation tasks

ToF cameras have been used to sense relatively large depth values for mapping or obstacle avoidance in mobile robotics, and also for human detection and interaction. At closer distances, ToF cameras have been applied to object modeling [5, 6], precise surface reconstruction [7], and to grasp known [8] and unknown [9] objects. We focus our review on two complementary areas: *scene-related tasks* and *object-related tasks*. *Scene-related tasks* generally involve moving the camera using a mobile robot. Although the range of distances involved is rather long, the techniques and ideas can be applied to eye-in-hand algorithms. *Object-related tasks* involve the use of ToF cameras at close distances. The most common application is object modeling, and to a lesser extent to enable object manipulation.

A table is provided in each section to summarize and give a comprehensive view of its contents. Our conclusion is that the most exploited feature of ToF cameras is their capability of delivering complete scene depth maps at high frame rate without the need of moving parts.

The depth-intensity image pair is also often used because both images are registered. In applications where high resolution is required a common solution is to fuse with color cameras. ToF cameras are used in human environments because they are eye-safe and permit avoiding physical contact and dedicated markers or hardware.

The first works that appeared were comparisons between ToF and other technologies. Then, in subsequent works, these technologies were gradually complemented, and sometimes substituted, by ToF sensors.

### *3.1. Scene-related tasks*

This kind of applications deal with tasks involving scenes that contain objects like furniture and walls. Observe that the expected range of distances to these objects is relatively wide. A usual framework in these applications is to install the camera on a mobile robot and use it for robot navigation and mapping. As it will be seen, one of the areas where ToF sensors are adequate is in obstacle avoidance, because the detection region is not only horizontal (like in laser scanners) but also vertical, allowing the robot to detect obstacles with complex shapes. Clearly, *the most appreciated characteristic of ToF sensors here is the high frame rate* (see Table 2). Some applications also benefit from the metric information obtained with depth images.

**Comparison**. Initial works were devoted to the comparison of ToF with other sensors, mainly laser scanners. Thanks to the larger vertical field of view of ToF cameras, difficult obstacles (like tables) are better detected by them than by 2D laser scanners. For example, Weingarten *et al.* [10] demonstrated this in the context of an obstacle avoidance algorithm.

To obtain a comparable detection area, a 3D scanner can be built from a pivoted 2D laser scanner. May *et al.* [11, 12] compared the performance of their robot navigation algorithm using such sensor and using a ToF camera. One of the main difficulties they encountered is the accumulated error in the map created with the ToF camera, leading to failures when closing loops, for instance. Compared to pivoted laser scanners, accumulated errors usually occur more often with ToF cameras due to their smaller field of view. As we will see in the next section, this problem is also present in objects modeling tasks.

**Only ToF**. ToF sensors have been used successfully as the unique sensor in some mobile robotic applications, despite their characteristic limited resolution. For mapping purposes, *ToF sensors are very interesting because they allow to extract geometric features*. Most of the reviewed applications extract planar regions using both intensity and depth images. In [13], May *et al.* explored different methods to improve pose estimation. They propose additionally a final refinement step that involves the alignment of corresponding surface normals leading to improved 3D scene maps computed at frame rate. The normal of the extracted planes is also used by Hedge and Ye [14] to detect badly conditioned plane detection, as horizontal planes in a staircase. Also Pathak *et al.* [15] have reported the use of ToF to extract planes for 3D mapping.

Alternatively, the acquired crude point clouds can be processed by a variant of the Iterative Closest Point (ICP) algorithm to find the relation between two point clouds. For example, a real time 3D map construction algorithm is proposed by Ohno *et al.* [16] in the context of a snake-like rescue robot operating in complex environments, like rubble in disaster-like scenarios. Here, a modification of the classical ICP algorithm is proposed to cope with ToF noisy readings and to speed up the process.

Another adaptation of an ICP-like algorithm for ToF images is presented by Stipes *et al.* [17], where both the depth and the intensity images are used. They present a probabilistic point sampling process to obtain significant points used in the registration process.

ICP assumes that both point clouds overlap, so wrong depth points can distort the result. May *et al.* [18] presented an ICP variant to take this explicitly into account. They propose a mapping algorithm using a Simultaneous Localization and Mapping (SLAM) technique to reduce the reconstruction error that is specially useful when a zone of the scenario is revisited, i. e., when closing a loop.

Also with potential applications to SLAM, Gemeiner *et al.* [19] proposed a corner filtering scheme combining both the intensity and depth images of a ToF camera.

Complex environments are a good test field for ToF sensors, as they are capable of naturally recovering their geometry. In the context of pipeline inspection, Thielemann *et al.* [20] have proposed to use a ToF camera to detect the different junctions based not on appearance but on geometric properties. Here the self-illumination mechanism of ToF sensors is appreciated. Furthermore, Sheh *et al.* [21] have proposed a ToF based navigation system for a random stepfield terrain[3]. They use the depth information to color an array of pixels and then perform some classical edge detection algorithms in this array, which is called *heightfield*. The heading and attitude compensation of the image is performed using an inertial unit.

ToF sensors have proved to be also applicable in dynamic environment mapping thanks to their characteristic high frame rate. Swadzba *et al.* [22] present a scene reconstruction algorithm that discards dynamic objects, like pedestrians, using a static camera in the difficult case of short sequences (2-3 sec.). Motion is recovered via optical flow in the intensity images, and then transferred to the depth image to compute a 3D velocity vector.

ToF cameras have been employed also in the automotive field to assist in parking operations. In [23] Acharya *et al.* describe the system design of a ToF camera for backup obstacle detection. In [24] the same group presents an application of a similar camera for the detection of curves and ramps also in parking settings. A modified Ransac algorithm, that uses only the best inliers, is used to find the best fitting of the planar patches that model the environment. ToF has been used also to control the deployment of the airbag system depending on the nature of the occupant in a car [25]: adult, child, child seat or objects.

**Fusion with other sensors**. Some authors have started recently to fuse ToF cameras with other sensors, i.e. laser scanners and different types of color cameras. A simple approach is to integrate ToF into existing algorithms. For example, Yuan *et al.* [26] propose a fusion process to integrate 3D data in the domain of laser data by projecting ToF point clouds onto the laser plane. This is applicable when considering a simple shaped robot, i.e. one that can be approximated by a cylinder, and it

[3] Stepfield terrains are the NIST proposal to generate repeatable terrain for evaluating robot mobility.

Table 2: ToF camera usage in scene-related tasks

| Article | Topic | Advantages | Type of Sensor |
|---|---|---|---|
| Weingarten *et al.* [10] | Obstacle avoidance in static env. | 3D at high rate | SR2 (depth) |
| May *et al.* [11, 12] | 3D mapping | 3D at high rate/No required Pan-Tilt | SR2 (depth) |
| May *et al.* [13] | Pose estimation/3D mapping | Registered depth-intensity | SR3 (depth + intensity) |
| Hedge and Ye [14] | Planar feature 3D mapping | 3D at high rate/No required Pan-Tilt | SR3 |
| Ohno *et al.* [16] | 3D mapping | 3D at high rate | SR2 |
| Stipes *et al.* [17] | 3D mapping / Point selection | Registered depth-intensity | SR3 |
| May *et al.* [18] | 3D mapping/SLAM | 3D at high rate | SR3 |
| Gemeiner *et al.* [19] | Corner filtering | Registered depth-intensity | SR3 (depth + intensity) |
| Thielemann *et al.* [20] | Navigation in pipelines | 3D allow geometric primitives search | SR3 |
| Sheh *et al.* [21] | Navigation in hard env. | 3D at high rate | SR3 + inertial |
| Swadzba *et al.* [22] | 3D mapping in dynamic env. | 3D at high rate/Registered depth-intensity | SR3 (depth + intensity) |
| Acharya *et al.* [23]<br>Gallo *et al.* [24] | Safe car parking | Improved depth range/3D at high rate | Canesta |
| Gortuk *et al.* [25] | Object classification (airbag app.) | light/texture/shadow independence | Canesta |
| Yuan *et al.* [26] | Navigation and obst. avoidance | Increased detection zone | SR3 + laser |
| Kuhnert and Stommel *et al.* [27] | 3D reconstruction | Easy color registration | PMD + stereo |
| Netramai *et al.* [28] | Motion estimation | 3D at high rate | PMD + stereo |
| Huhle *et al.* [29] | 3D mapping | Easy registration of depth and color | PMD + color camera |
| Prusak et al. [30] | Obst. avoidance/Map building | Absolute scale/better pose estimation | PMD + spherical camera |
| Swadzba *et al.* [31] | 3D mapping/Map optimization | 3D at high rate | SR3 |
| Vaskevicius *et al.* [32]<br>Poppinga [33] | Localization/Map optimisation | Neighbourhood relation of pixels<br>No color restrictions | SR3 |

entails a minimum update of their previous laser-scanner-based algorithm. Nevertheless, the resulting algorithm can cope with new kinds of obstacles in a simple way. Note that this is not a pure 3D approach and it is not using the potentiality of having full 3D information at a high frame rate.

Fusion of color and depth information in scene tasks seems to have a great potential. In a preliminary work, Kuhnert and Stommel [27] present a revision of their 3D environment reconstruction algorithm combining information from a stereo system and a ToF sensor. Later, Netramai *et al.* [28] compared the performance of a motion estimation algorithm using both ToF and depth from stereo. They also presented an oversimplified fusion algorithm that relies on the optical calibration of both sensors to solve the correspondence problem. These works propose fusion paradigms combining the results produced in two almost independent processes.

Contrarily, Huhle *et al.* [29] present a color-ICP algorithm useful for scene-based image registration, showing that introducing color information from a classical camera in the beginning of the process effectively increases the registration quality.

Depth information allows to identify in a robust manner not only obstacles but also holes and depressions. Prusak *et al.* [30] proposed a join approach to pose estimation, map building, robot navigation and collision avoidance. The authors use a PMD camera combined with a high-resolution spherical camera in order to exploit both the wide field of view of the latter for feature tracking and pose estimation, and the absolute scale of the former. The authors relied on a previous work on integration of 2D and 3D sensors [34, 35], showing how restrictions of standard Structure-from-Motion approaches (mainly scale ambiguity and the need for lateral movement) could be overcome by using a 3D range camera. The approach produced 3D maps in real-time, up to 3 frames per second, with an ICP-like algorithm and an incremental mapping approach.

**Noisy data enhancement**. Swadzba *et al.* [31] propose a new algorithm to cluster redundant points using a virtual plane, which apparently performs better in planar regions and reduces noise, improving registration results. Furthermore, a group at Jacobs University [32, 33] has proposed to identify surfaces using a region growing approach that allows the poligonization of the resulting regions in an incremental manner. The nature of the information delivered by ToF cameras, specially the neighborhood relation of the different points, is explicitly exploited and also their noisy nature is taken into account. Moreover, some comparisons with results from stereo rigs are reported.

Finally, Huhle *et al.* [36] propose an alternative representation of the map by means of the Normal Distribution Transform, which efficiently compresses the scan data reducing memory requirements. This representation seems to be well suited also for the typical noisy ToF depth images.

### *3.2. Object-related tasks*

ToF cameras have also been successfully used for object and small surface reconstruction, where the range of distances is small. A comprehensive summary is given in Table 3

**Comparison with stereovision**. A classical solution in the area of object modeling is the use of calibrated stereo rigs. Therefore, initial works were devoted to their comparison with

Table 3: ToF camera usage in object-related tasks

| Reference | Topic | Advantages | Type of Sensor |
|---|---|---|---|
| Ghobadi *et al.* [37] | Dynamic object detection and classification | Color and light independence | PMD |
| Hussmann and Liepert [38] | Object pose | Easy object/background segmentation | PMD |
| Guomundsson *et al.* [39] | Known object pose estimation | Light independent / Absolute scale | SR3 |
| Beder *et al.* [40] | Surface reconstruction using patchlets | ToF easily combines with stereo | PMD |
| Fuchs and May [7] | Precise surface reconstruction | 3D at high rate | SR3/O3D100 (Depth) |
| Dellen *et al.* [5]<br>Foix *et al.* [6] | 3D object reconstruction | 3D at high rate | SR3 (Depth) |
| Kuehnle *et al.* [8] | Object recognition for grasping | 3D allow geometric primitives search | SR3 |
| Grundmann *et al.* [41] | Collision free object manipulation | 3D at high rate | SR3 + stereo |
| Reiser and Kubacki [42] | Position based visual servoing | 3D is simply obtained / No model needed | SR3 (Depth) |
| Gachter *et al.* [43]<br>Shin *et al.* [44] | Object part detection for classification | 3D at high rate | SR3<br>SR2 |
| Klank *et al.* [45] | Mobile manipulation | Easy table/object segmentation | SR4 |
| Marton *et al.* [46] | Object categorization | ToF easily combines with stereo | SR4 + color |
| Nakamura *et al.* [47] | Mobile manipulation | Easy table segmentation | SR4 + color |
| Saxena *et al.* [9] | Grasping unknown objects | 3D at high rate | SR3 + stereo |
| Zhu *et al.* [48] | Short range depth maps | ToF easily combines with stereo | SR3 + stereo |
| Lindner *et al.* [49] | Object segmentation for recognition | Easy color registration | PMD + color camera |
| Fischer *et al.* [50] | Occlusion handling in virtual objects | 3D at high rate | PMD + color camera |

ToF sensors showing the potential of the latter when poorly textured objects are considered, and when background-foreground segmentation is difficult. For planar and untextured object surfaces, where stereo techniques clearly fail, Ghobadi *et al.* [37] compared the results of a dynamic object detection algorithm based on SVM using stereo and ToF depth images. In the same manner, Hussmann and Liepert [38] also compared ToF and stereo vision for object pose computation. The key difference favorable to ToF camera is its ability to effectively segment the object and the background, even if their color or texture is exactly the same (i.e. a white object on a white table). They also propose a simple method to obtain object pose from a depth image.

Another comparison is presented by Guomundsson *et al.* [39]. They classify and estimate the pose of some simple geometric objects using a Local Linear Embedding (LLE) algorithm, and contrast the results of using the intensity image and the depth image. Their analysis shows that range data adds robustness to the model, simplifies some preprocessing steps, and in general the generated models capture better the nature of the object. Stereo and ToF have also been compared by Beder *et al.* [40] in the framework of surface patchlet identification and pose estimation. In their setup, using a highly textured surface for stereo experiments, ToF slightly outperforms stereo in terms of depth and normal direction to the patchlet. Thus, ToF can be used to benchmark stereo surface reconstruction algorithms.

**ToF for surface reconstruction**. To obtain 3D object surfaces, multiple 3D images need to be acquired and the resulting 3D point clouds should be combined. The setups for these object modeling algorithms usually include a ToF camera mounted on the end-effector of a robotic arm. *Point cloud registration is more critical in object modeling than in scene modeling*. Even if the hand-eye system is precisely calibrated, the displacement given by the robot is usually not enough and the transformation between different point clouds has to be calculated. The application of ICP in two consecutive views naturally accumulates errors and consequently more precise algorithms need to be used.

To obtain precise object models, Fuchs and May [7] perform a circular trajectory around the object to acquire equally spaced images, and use a *simultaneous matching* algorithm [51] instead of classical ICP to distribute the errors in all the estimated displacements. Their work also includes a comparison of two different ToF cameras. Alternatively, Dellen *et al.* [5] propose a fine registration algorithm based on an ICP algorithm using invariant geometric features. The resulting model is obtained after reducing noise and outliers by treating the coarse registered point cloud as a system of interacting masses connected via elastic forces. Alternatively, Foix et al. [6] propose a method to compute the covariance of the point clouds registration process (ICP), and apply an iterative view-based aggregation method to build object models under noisy conditions. Their method does not need accurate hand-eye calibration since it uses globally consistent probabilistic data fusion by means of a view-based information-form SLAM algorithm, and can be executed in real time taking full advantage of the high frame rate of the ToF camera.

**ToF for object manipulation**. Object recognition and object pose estimation algorithms are usually related to robotic manipulation applications: objects have to be identified or categorized with the aim of finding and extracting some characteristics to interact with them. This is usually a challenging task as ToF depth images are noisy, and low sensor resolution leads to only few depth points per object.

Kuehnle *et al.* [8] explore the use a ToF camera to recognize and locate 3D objects in the framework of the robotic manipulation system DESIRE. Objects are modeled with geometric primitives. Although they use depth images rectified up to some level, their system is not reliable enough. In a subsequent work [41] they use the ToF camera to detect unknown objects and classify them as obstacles, and use a stereo camera system to identify known objects using SIFT features. As it is widely known, this second approach requires textured objects while their first approach does not. In the same project, Reiser and Kubacki [42] have proposed a method to actively orientate the camera using a visual servoing approach to control a pan-and-tilt unit. They proved that position-based visual servoing is straightforward by using a ToF camera, because of its ability to deliver 3D images at high rate.

In a different way, Gächter *et al.* [43] propose to detect and classify objects by identifying their different parts. For example, chairs are modeled by finding their legs, which in turn are modeled with vertical bounding boxes. The tracking of the different parts in the image sequence is performed using an extended particle filter, and the recognition algorithm is based on a SVM, that proves again to be useful in typical noisy ToF images. Later, Shin *et al.* [44] used this incremental part detector to propose a classification algorithm based on a geometric grammar. However, they use a simulated environment because the classification in real scenarios does not seem to be reliable enough.

ToF cameras have been also used in the framework of mobile manipulation, where a mobile robot has the task to detect and grasp unknown objects. Depth information here is very useful in both clean and cluttered environments. Klank *et al.* [45] propose a mobile manipulation algorithm where a ToF camera is mounted on the end-effector of a robot arm embarked on a mobile robot. An in eye-in-hand configuration provides a large mobility to the camera, allowing to easily change the point of view. The assumption is that objects would be on top of supporting planes, e.g. a table. In their algorithm, once the table has been located, the corresponding 3D points are removed from the image. The remaining points would correspond to objects.

One advantage of ToF cameras is that the 3D region of interest can be extracted easily. Another advantage is that some object segmentation algorithms can be developed combining cues from both a ToF sensor and a color camera. Using such a combined sensor, Marton *et al.* [46] proposed a probabilistic categorization algorithm for kitchen objects. This work uses a new SR4000 camera. This sensor assigns a confidence value to each depth reading that allows to infer if the object material is producing bad sensor readings.

Combining the two last ideas, that is, table plane extraction and depth-color combination, Nakamura *et al.* [47] propose an algorithm to move a mobile robot, with a ToF camera and two CDD cameras mounted on its head, next to the supporting table where objects are supposed to be. Their proposal uses the depth to easily remove the 3D points corresponding to the table, and to cluster the remaining points. Color is used then at each of the clusters to recognize objects.

Thanks to the depth information, some grasping properties can be easier to evaluate, i.e. form- and force-closure, sufficient contact with the object, distance to obstacles, and distance between the center of the object and the contact point. Saxena *et al.* [9] used this advantage to propose a learning grasp strategy that identifies good grasping points using partial shape information of unknown objects. The contribution of the depth information allows to update an already presented method using a color camera, with the advantage of having depths even in texture-less portions of the objects.

**Fusion algorithms**. In fact, ToF and stereo systems naturally complement one another. As has been argued before, ToF performs correctly in poorly textured surfaces and object segmentation becomes easy even in poorly contrasted situations. Contrarily, it has difficulties precisely in textured surfaces and in short distances, where stereo outperforms it. This fact has been exploited in several works. For example, Zhu *et al.* [48] propose a probabilistic framework to fuse depth maps from stereo and the ToF sensor. They use a depth calibration method to improve the ToF image, which is useful in small depth ranges (from 1m to 1.4m).

Another fusion framework is proposed by Lindner *et al.* [49] using calibration and scaling algorithms. They obtain a dense colored depth map using the geometrical points correspondence between the ToF and color cameras by assigning a color to the ToF depth points, and interpolating the depth of the rest of the color camera pixels. A way to detect areas not seen by the color camera is also provided, as well as some techniques to enhance edges and detect invalid pixels.

Finally, in the context of augmented reality, Fischer *et al.* [50] combine a ToF camera and a standard color camera to handle virtual object occlusions caused by real objects in the scene. Fast 3D information is highly valuable, as well as its independence on lightning conditions, object texture and color. They do not use any depth calibration or noise outliers removal algorithm, and consequently the negative effect of noise is clearly visible in their results.

**Summary and final remarks**.

ToF cameras have been successfully used for object and small surface reconstruction at close distances. In general the scenario for these applications involves a robotic manipulator and the task requires modeling object shape. In such settings, one has to expect that some over-saturation problems may occur when acquiring depth images. On the contrary, as the range of depths is short, calibration can be simplified.

Some of the reviewed works do not apply any calibration method to rectify the depth images. We believe that this explains some of the errors and inaccuracies reported in some experiments, and that with proper calibration better results can be obtained. We note that ToF technology is evolving and depth correction methods are still subject to investigation.

Foreground/background segmentation methods based on depth information are quite straightforward, so ToF images are used in many applications requiring them. A good characteristic is that geometric invariants as well as metric constraints can be naturally used with the ToF depth images.

ICP-like techniques are the preferred solution to reconstruct

surfaces. A common approach to identify objects is the use of Support Vector Machines, which perform adequately when considering the noisy point models obtained with one ToF image or when merging different ToF views.

The high frame rate of ToF sensors is a key advantage, but also the natural combination with color cameras and stereo rigs. The fact that the depth and intensity images are delivered already registered is handy in some contexts, but in applications where the reduced resolution of a ToF camera is critical, it is complemented with other sensors, usually color cameras. Actually, a growing trend is observed not to use the intensity image supplied by the ToF camera, preferring the combination with high-resolution conventional cameras.

## 4. Depth measurement errors and calibration

Raw measurements captured by ToF cameras provide noisy depth data. Default factory calibration can be used in some applications where accuracy is not a strong requirement and the allowed depth range is very large. For the rest of applications ToF cameras have to be specifically calibrated over the defined application depth range. Two types of errors, *systematic* and *non-systematic*, can interfere and consequently corrupt ToF depth readings. A detailed ToF error description and classification can be found in [2].

ToF cameras are evolving and a lot of work is being carried out to understand the source of errors and to compensate them. For example, recent works demonstrate the benefits of using a band-pass optical filter to increase the sensor performance and also attenuate the background light [52].

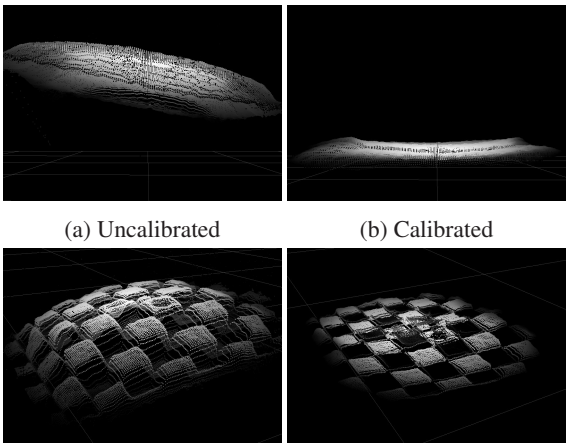
### *4.1. Systematic errors*

ToF camera systematic errors can be either constant or dependent on the actual measurement. The advantage of systematic errors is that depth values can be corrected through calibration. Two of the most important systematic errors are *depth distortion*, an offset following a sinusoidal shape depending on the measured depth that affects all the image (compare the height (depth) of the plane in Fig. 3a and 3b), and *built-in pixel errors*, which is a constant offset of each pixel independent of the measured depth. It can be observed as a rotation of the whole scene (compare the orientation of the plane in Fig. 3a and 3b).

A very important parameter to control in ToF cameras is the integration time (IT). Simplifying, IT can be seen as the shutter-time parameter in conventional cameras. It is observed that measurements performed with a static camera while changing only the IT value produce different depth results. This is called the *integration time error*, and it is usually solved by calibrating the camera at each one of the different used IT values.

Observing Fig. 3c, it can be seen that a planar surface is perceived as non-planar. Depth at the borders, where illumination is less intense, is over-estimated. In contrast, if the object is too close (or IT too high) the saturation leads to underestimation of the depth. This is caused by the *amplitude-related errors* and can be corrected by means of calibration (Fig. 3d).

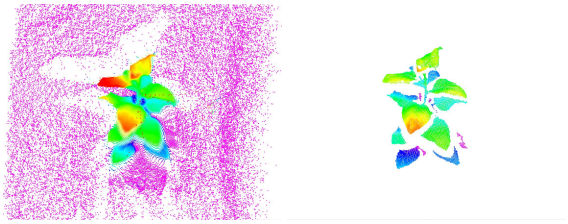
The last systematic error to take into account is related to temperature. The measurements drift until the camera reaches the working temperature. The usual solution is to wait some time until the temperature is stable.



(a) Uncalibrated (b) Calibrated

(c) Uncalibrated (d) Calibrated

Figure 3: 3D image of a planar surface and a white/black checker-board. (a) surface should be horizontal, but *built-in pixel-related* error causes a distortion. (b) Once calibrated, the orientation of the plane and the depth of individual points is corrected. (c) Observe the difference in depth between the squares of each color. (d) The calibrated image is rectified taking into account built-in and amplitude errors.



(a) Raw image (b) Filtered

Figure 4: Reduction of noise by filtering pixels using a flying-points detector and depth threshold filtering.

### *4.2. Non-systematic errors*

Non-systematic errors appear because of unknown and unpredictable situations. Contrary to systematic errors, pixels with wrong depth values cannot be recovered and have to be removed. Newer cameras can detect and mark invalid pixels. The two main reasons are bad depth estimation due to darkish pixels (very low illumination) or over-exposed pixels (too much illumination). Filtering algorithms can be applied to minimize the remaining bad pixels.

The most important error is due to *multiple light reception*. This causes one of the know problems with ToF cameras, the so called *flying points*. These are false points that appear between the edges of the objects and the background. These points can be easily located in the depth image and the 3D point cloud, and easy-to-implement filtering methods are available [2].

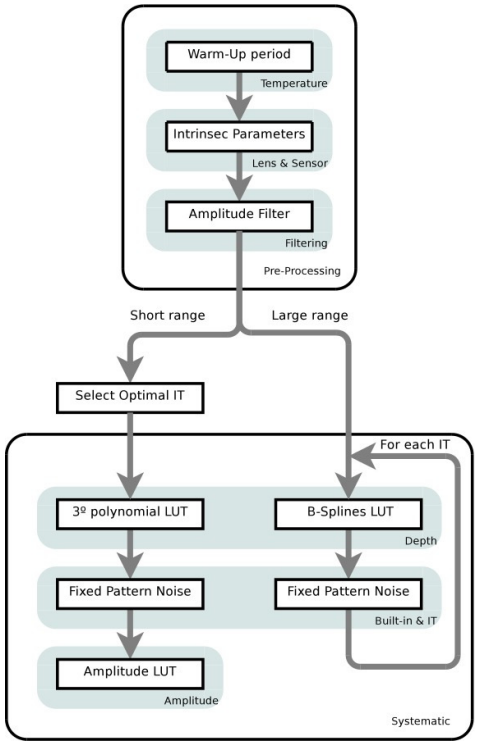



Figure 5: Suggested calibration approach dependent on the range of the scene. Short-range (0.4m - 2m). Large-range (1.5m - 7.5m).

### 4.3. Depth calibration

Several calibration methods have been proposed, each one with its own peculiarities. Among them, it is quite difficult to decide which calibration method one should apply in each situation. Here we will provide a general unified approach to camera calibration.

A naive such approach would be to parameterize per pixel the complete set of systematic errors. Applying such an approach would require a 4-D look-up table providing error offsets dependent on pixel position, integration time, measured depth and reflected amplitude. Although the naive calibration would be very accurate, it would be highly time consuming and consequently inadvisable for robotic applications.

A general approach consisting of two calibration processes can be devised by distinguishing between short-range and long-range robotic applications. Figure 5 is a graphical representation of this approach. Short-range applications are defined as the ones whose depth ranges between 0.4 and 2 meters, and long-range applications are those with a depth range larger than 2 meters and which do not have measurements closer than 1.5 meters. While the pre-processing module is shared, the systematic errors calibration module is tailored to each process. Large-range applications differ from short-range ones in that the scene can contain simultaneously nearby and distant objects. That is the reason why multiple IT look-up tables have to be built and B-Splines are used, instead of the simple polynomials used in the short range. In the same manner, multiple FPN are calculated for each IT, while short-range applications need only one. Systematic amplitude errors due to object reflectivity can be considered negligible for large range scenes if compared to the other influencing factors [54, 53].

**Algorithm 1** Suggested Calibration Approach

```
wait warm-up period;
calculate intrinsic parameters;
filter amplitude;
if ShortRangeScene then
    Select Optimal Integration Time;
    LUT_IT = compute 3-degree polynomial look-up table;
    FPN = compute fixed pattern noise;
    LUT_A = compute amplitude look-up table;
else
    N_i = choose integration time set;
    for all i such that 1 ≤ i ≤ size(N_i) do
        LUT_i = compute B-Spline look-up table for N_i;
        FPN_i = compute fixed pattern noise for N_i;
    end for
end if
```

Instead of solving each systematic error step by step, a complete calibration approach can be used for short range applications.

The pre-processing module is divided into three stages. First of all, a *warm-up period* has to be waited in order to minimize temperature effects, usually between 10 and 20 minutes depending on the camera. Secondly, *intrinsic parameters* have to be calculated in order to convert from spherical coordinates to Cartesian ones taking into account lens and sensor distortions. And finally, an amplitude filtering is applied in order to increase accuracy by using more reliable data. Low illuminated and over-saturated pixels have to be discarded before a systematic error calibration process takes place.

Short range applications have the peculiarity of requiring only a particular integration time for covering the whole scene. The optimal IT can be chosen as the one with less over-saturated and underexposed pixels at the required depth range. Once the IT is defined, systematic errors can be estimated. Firstly, a 3-degree polynomial look-up table is built by measuring depth of a mobile white planar surface at several intervals over the defined short range. Only a centered pixel is used for the measurements. Afterwards, a fixed pattern noise is computed at half depth range, this time for the whole sensor matrix. And finally, an amplitude look-up table is built by means of a mobile scaled black-to-white calibration pattern, once again using a single centered pixel.

Large range applications need different integration times depending on the current scene distance. That's the reason why, after choosing the required IT range, multiple IT look-up tables have to be calculated. Considering that the scene can contain different distant objects simultaneously, each IT look-up table has to be measured over the full range in order to evaluate all measurements. B-Splines are used, instead of a polynomial, in order to reduce the overall complexity. Measurements at a

central pixel of a mobile planar surface are used here too.

## 5. Applications

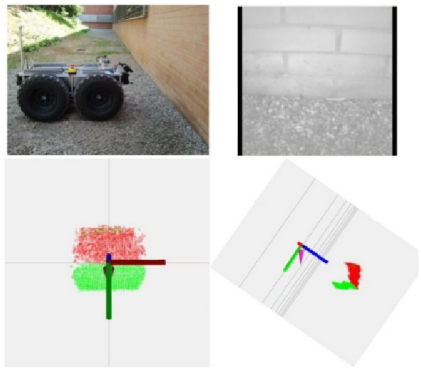
Some use-cases of ToF cameras in scene- and object-related tasks are presented in this section. In both contexts the position of the camera can be actively changed. However, when the camera is mounted on a mobile robot the set of feasible new camera positions is restricted (Sec. 5.1). Clearly, as will be shown in the object-related subsection (Sec. 5.2), the eye-in-hand configuration provides a greater mobility of the camera.

### *5.1. Scene-related tasks*

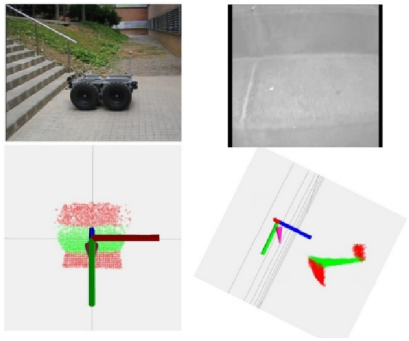
Scene-related tasks usually involve a ToF camera mounted on a mobile robot. The placement of the camera highly depends on the task. When the task is navigation and mapping, the camera is placed horizontal on the front of the vehicle. Compared to the common laser solution, the advantage of ToF cameras here is that they provide a full image at frame-rate that can be used, for example, to easily detect difficult obstacles. Contrarily, when the task is mobile manipulation, for example a robot that navigates in the environment to locate and grasp a given object, the camera is placed at some height, usually at the 'head' of the robot, looking downwards to the working area of the manipulator arm. This is sometimes called the hand-eye configuration, opposed to the eye-in-hand configuration described in the next section.

We present an example of an all-terrain robot that uses a PMD Camcube ToF camera, placed in the front and facing downwards, for robust obstacle detection in real-time outdoor mobile navigation. The main difficulties for most sensors outdoors are caused by sunlight effects. However, it has been demonstrated that ToF cameras can be configured to deliver correct enough images in sunlight [4].

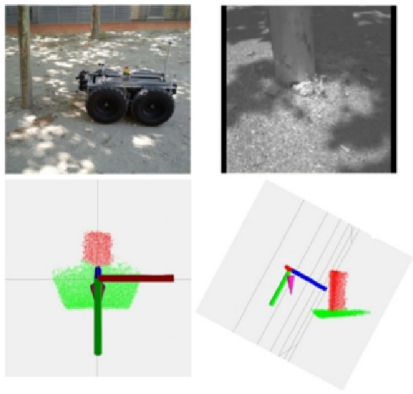
Figure 6 shows the robot facing three different outdoor obstacles: a wall (Fig. 6a), stairs (Fig. 6b) and a tree trunk (Fig. 6c). Observe in the figures that depth data is much more informative than intensity. The segmentation of 3D structures, like the bottom edge of a wall, is more robust using depth information. The algorithm is based on the computation of normals for each point, and on the comparison of neighboring normals to form surfaces. Authors demonstrate that with this algorithm holes in the terrain can be also correctly detected.



(a) Wall.



(b) Stairs.



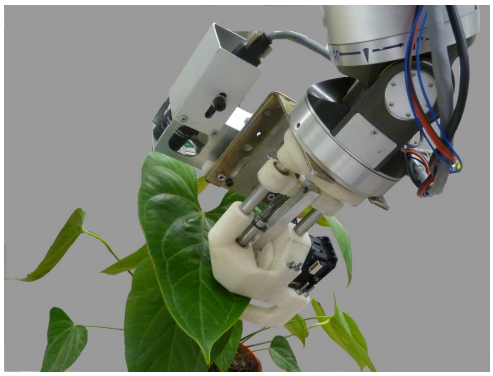
(c) Tree.

Figure 6: Mobile robot using a ToF camera for obstacle detection in three different outdoor scenarios. In each figure can be observed: (1)robot scene, (2)ToF intensity image, (3)ToF pointcloud segmented (green points represent horizontal planes, red points vertical planes), (4)ToF pointcloud rotated. Reproduced from [1].

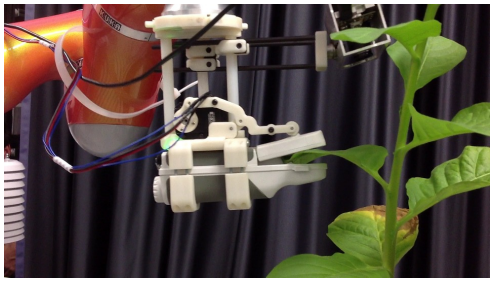
### *5.2. Object-related tasks*

Two of the main advantages of actively changing the point of view of a ToF camera are highlighted: the easy acquisition of 3D structure (that allows straightforward foreground-background segmentation), and the ability to acquire accurate views of particular details of the scene.

The examples are based on recent experiences mainly in the field of plant phenotyping, and to a lesser extent in that of textile manipulation. In plant phenotyping, a large number of plants has to be monitored searching for unusual plant responses to external factors as extreme humidity or poor watering. Nowadays, automation of greenhouses provides automatic conveyor belts

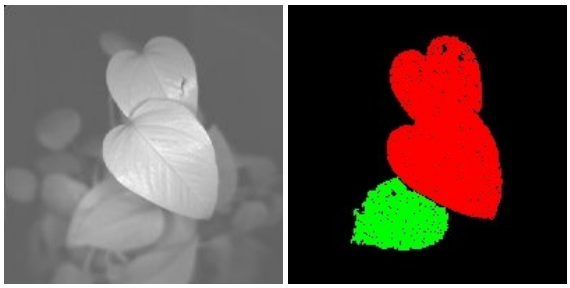


(a) Custom cutting tool and ToF-color camera set.

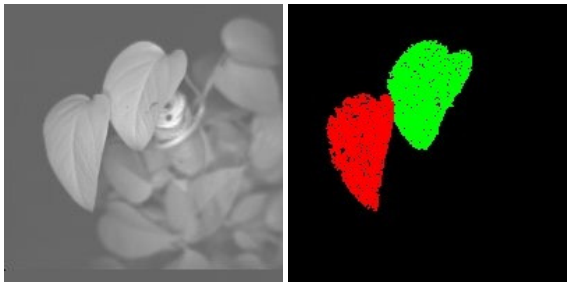


(b) Chlorophyll meter and ToF camera.

Figure 7: Details of two different tools in the end-effector of (a) a WAM robot and (b) a Kuka Lightweight robot. Both tools require that the leaf is placed inside their lateral aperture. An eye-in-hand ToF camera permits acquiring the 3D plant structure required to compute robot motion.



(a) Frame 1: intensity image and segmentation



(b) Frame 2: intensity image and segmentation

Figure 8: Scene containing three leaves initially identified as two segments (Frame 1). After changing the point of view (Frame 2), the ambiguity is cleared and two leaves are detected that were joined in one segment in the first frame.. Additionally, the whole surface of the initially partial-occluded leaf is now visible. Observe that the effectiveness of the camera motion depends on the particular leaf arrangement.

to transport plants to a measuring cabin, where a set of sensors performs all the measurements required. However, plants can have complex shapes, and having to define the best static position for all the cameras and other sensors is problematic. The ability to mount a sensor on a manipulator robot in an eye-in-hand configuration is highly appreciated. Additionally, some tasks require to place the sensor or the tool on the surface of a leaf. We provide here two examples of such tasks: the measurement of chlorophyll with a SpadMeter, and the extraction of sample discs for DNA analysis (see in Fig. 7 both scenarios with the ToF cameras in an eye-in-hand configuration).

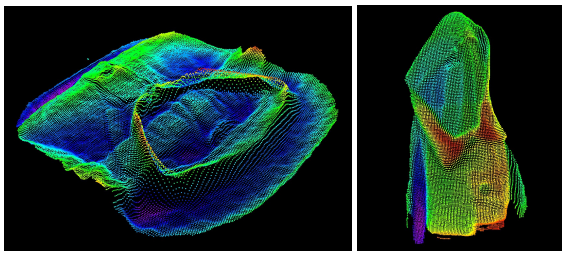
*3D structure and disambiguation.* One of the objectives in plant phenotyping is to gather as much information as possible about each specimen, preferably 3D relevant information to enable subsequent manipulation. Color vision is helpful to extract some relevant features, but it is not well-suited for providing the structural/geometric information indispensable for robot interaction with plants. 3D cameras are, thus, a good complement, since they directly provide depth images [55]. Moreover, plant data acquired from a given viewpoint are often partial due to self-occlusions, thus planning the best next viewpoint becomes an important requirement. This, together with the need of a high throughput imposed by the application, makes 3D cameras (which provide images at more than 25 frames-per-second) a good option in front of other depth measuring procedures, such as stereovision or laser scanners.

Segmentation algorithms use different parameters that need to be adapted to the characteristics of the data, like long ranges, noise type, and sensitivity. The eye-in-hand approach permits moving the camera to find the view that fits better the segmentation parameters. Figure 8 shows an example, where in the first view the segmentation algorithm, that uses depth similarity between adjusted surfaces, fails to distinguish two different leaves. Using a next-best-view algorithm [6], a new view is selected that maximizes the difference in depth of the two leaves, thus the algorithm is now capable of distinguishing the two leaves.

Additionally, the partial occlusion of the rear leaf produced by the front leaf (Fig. 8a) is resolved after the motion of the camera (Fig. 8b). The benefits of moving the camera have some limits in such complex scenarios, as it is not always possible to obtain a better viewpoint, for example when occlusions are too strong, or when the optimal point of view is out of the working space of the robot.

*Detailed views.* The eye-in-hand configuration allows to control not only the viewpoint of the camera, but also the distance to the object. To change the distance is also an strategy to change the effective resolution of the image, as relevant details can be better focused.

Figure 9 shows the image of a shirt in two different configurations: folded and hanged. Here the task is to grasp the shirt



(a) Folded (b) Hanged

Figure 9: Details of the perception of a shirt in different configurations. Observe that the small wrinkles are correctly perceived, and some characteristic parts, like the collar shape, are clearly visible. Depth is codified as color.

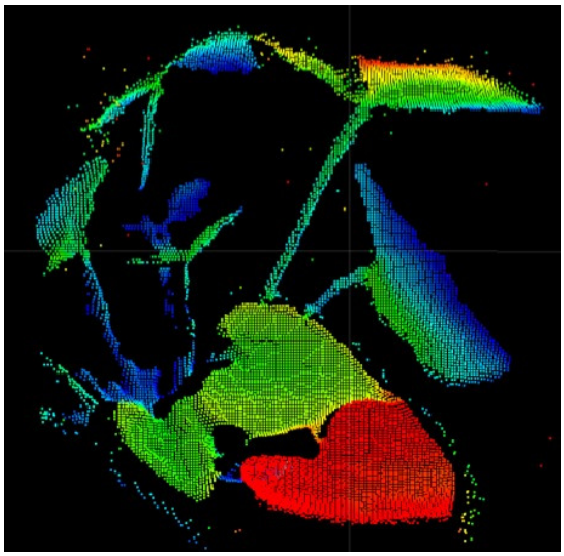


Figure 10: Detail of a plant. Observe that the stems, even if they are thin, are correctly acquired.

from the collar to allow the robot to hang the shirt in a hanger. Observe that in both configurations the details of the collar, the buttons and small winkles are visible. In the hanged shirt the sleeves are identifiable as well. Previous works have shown that this 3D structure can be used to identify wrinkles [56] and also the collar structure, using computer vision algorithms [57].

Clearly, the point of view determines the nature of the gathered information, but also the sensor sensitivity determines the relevant details that can be observed. Figure 10 shows a view of a plant where the stems are visible. Here, the point of view is important, but also that ToF cameras are sensible enough to capture these structures. This is hard to obtain with classical stereovision, and completely impossible with other sensors, like Kinect.

## 6. Conclusions

ToF cameras have been presented from different perspectives, including: underlying principle and characteristics, calibration techniques, applications where camera advantages are explicitly exploited, and potential for future research. Over the last years, performance of ToF cameras has improved significantly; errors have been minimized and higher resolution and frame rates have been obtained. Although ToF cameras cannot yet attain the depth accuracy offered by other types of sensors such as laser scanners, plenty of research demonstrates that they perform better in many robotic applications. The application of ToF cameras in the wide range of scientific areas we have reviewed indicates their great potential, and widens the horizon of possibilities that were envisaged in the past for vision-based robotics research.

Based on the task, we have divided the application fields in scene-related and object-related tasks. The former often involve a mobile robot and relatively long sensing distances. We have provided an example of a robot using ToF for obstacle avoidance and terrain classification. Object-related tasks involve a robotic arm, that confers great mobility to the camera, but reduces the distance to the viewed target. We have provided experimental evidence of the effectiveness of such approach in two tasks: plant 3D structure recovery and disambiguation, and the acquisition of detailed views.

Advantages of this type of sensors are multiple: they are compact and portable, easing movement; they make data extraction simpler and quicker, reducing power consumption and computational time; and they offer a combination of images that shows great potential in the development of data feature extraction, registration, reconstruction, planning and optimization algorithms, among other positive characteristics. Thus, ToF cameras prove to be especially adequate for eye-in-hand and real-time applications in general, and in particular for automatic acquisition of 3D models requiring sensor movement and on-line calculation.

Finally, some broad challenges need to be mentioned. First, resolution is still generally low for ToF cameras, despite some efforts have already led to better resolutions as explained above. Second, short integration times produce a high noise-to-signal ratio, and high integration times can result in pixel saturation [10]. Although some algorithms dealing with these problems have already been proposed, more research is needed in this direction. Third, the bi-static configuration (different positions of the emitter and the receiver) causes problems in close range situations because the measured intensity is sensitive to the varying illumination angle. The ability to move the camera is crucial to minimize this effect.


## Acknowledgements

This work was supported by the EU project GARNICS FP7-247947, by the Spanish Ministry of Science and Innovation under project PAU+ DPI2011-27510, and by the Catalan Research Commission through SGR-00155.



[1] À. Santamaria-Navarro, E.H. Teniente, M. Morta and J. Andrade-Cetto. Terrain classification in complex 3D outdoor environments. Journal of Field Robotics, 2014, to appear.

[2] S. Foix, G. Alenyà, C. Torras, Lock-in time-of-flight (ToF) cameras: A survey, IEEE Sensors J. 11 (9) (2011) 1917–1926.

[3] R. Lange, P. Seitz, Solid-state time-of-flight range camera, IEEE J. Quantum Electron. 37 (3) (2001) 390–397. doi:10.1109/3.910448.

[4] W. Kazmi, S. Foix, G. Alenyà and H.J. Andersen. Indoor and outdoor depth imaging of leaves with time-of-flight and stereo vision sensors: Analysis and comparison. ISPRS Journal of Photogrammetry and Remote Sensing, 88 (2014) 128–146.

[5] B. Dellen, G. Alenyà, S. Foix, C. Torras, 3D object reconstruction from Swissranger sensors data using a spring-mass model, in: Proc. 4th Int. Conf. Comput. Vis. Theory Appl., Vol. 2, Lisbon, 2009, pp. 368–372.

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

[7] S. Fuchs, S. May, Calibration and registration for precise surface reconstruction with time of flight cameras, Int. J. Int. Syst. Tech. App. 5 (3-4) (2008) 274–284. doi:http://dx.doi.org/10.1504/IJISTA.2008.021290.

[8] J. U. Kuehnle, Z. Xue, M. Stotz, J. M. Zoellner, A. Verl, R. Dillmann, Grasping in depth maps of time-of-flight cameras, in: Proc. Int. Workshop Robot. Sensors Environ., Otawa, 2008, pp. 132–137. doi:10.1109/ROSE.2008.4669194.

[9] A. Saxena, L. Wong, A. Y. Ng., Learning grasp strategies with partial shape information, in: Proc. 23th AAAI Conf. on Artificial Intelligence, Chicago, 2008, pp. 1491–1494.

[10] J. Weingarten, G. Gruener, R. Siegwart, A state-of-the-art 3D sensor for robot navigation, in: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Vol. 3, Sendei, 2004, pp. 2155–2160. doi:10.1109/IROS.2004.1389728.

[11] S. May, B. Werner, H. Surmann, K. Pervolz, 3D time-of-flight cameras for mobile robotics, in: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Beijing, 2006, pp. 790–795. doi:10.1109/IROS.2006.281670.

[12] S. May, K. Pervolz, H. Surmann, Vision systems: applications, I-Tech Education and Publishing, 2007, Ch. 3D cameras: 3D computer vision of wide scope, pp. 181–202.

[13] S. May, D. Droeschel, D. Holz, S. Fuchs, E. Malis, A. Nüchter, J. Hertzberg, Three-dimensional mapping with time-of-flight cameras, Journal of Field Robotics, Special Issue on Three-Dimensional Mapping, Part 2 26 (11-12) (2009) 934–965.

[14] G. Hedge, C. Ye, Extraction of planar features from Swissranger SR-3000 range images by a clustering method using normalized cuts, in: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Saint Louis, 2009, pp. 4034–4039. doi:10.1109/IROS.2009.5353952.

[15] K. Pathak, N. Vaskevicius, J. Poppinga, S. Schwertfeger, M. Pfingsthorn, A. Birk, Fast 3D mapping by matching planes extracted from range sensor point-clouds, in: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Saint Louis, 2009, pp. 1150–1155. doi:10.1109/IROS.2009.5354061.

[16] K. Ohno, T. Nomura, S. Tadokoro, Real-time robot trajectory estimation and 3D map construction using 3D camera, in: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Beijing, 2006, pp. 5279–5285. doi:10.1109/IROS.2006.282027.

[17] J. Stipes, J. Cole, J. Humphreys, 4D scan registration with the SR-3000 LIDAR, in: Proc. IEEE Int. Conf. Robotics Autom., Pasadena, 2008, pp. 2988–2993. doi:10.1109/ROBOT.2008.4543664.

[18] S. May, S. Fuchs, D. Droeschel, D. Holz, A. Nuechter, Robust 3D-mapping with time-of-flight cameras, in: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Saint Louis, 2009, pp. 1673–1678.

[19] P. Gemeiner, P. Jojic, M. Vincze, Selecting good corners for structure and motion recovery using a time-of-flight camera, in: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Saint Louis, 2009, pp. 5711–5716. doi:10.1109/IROS.2009.5354395.

[20] J. Thielemann, G. Breivik, A. Berge, Pipeline landmark detection for autonomous robot navigation using time-of-flight imagery, in: Proc. IEEE CVPR Workshops, Vol. 1-3, Anchorage, 2008, pp. 1572–1578. doi:10.1109/CVPRW.2008.4563167.

[21] R. Sheh, M. W. Kadous, C. Sammut, B. Hengst, Extracting terrain features from range images for autonomous random stepfield traversal, in: Proc. IEEE Int. Workshop Safety, Security Rescue Robot., Rome, 2007, pp. 24–29. doi:10.1109/SSRR.2007.4381260.

[22] A. Swadzba, N. Beuter, J. Schmidt, G. Sagerer, Tracking objects in 6D for reconstructing static scenes, in: Proc. IEEE CVPR Workshops, Vol. 1-3, Anchorage, 2008, pp. 1492–1498. doi:10.1109/CVPRW.2008.4563155.

[23] S. Acharya, C. Tracey, A. Rafii, System design of time-of-flight range camera for car park assistand backup application, in: Proc. IEEE CVPR Workshops, Vol. 1-3, Anchorage, 2008, pp. 1552–1557. doi:10.1109/CVPRW.2008.4563164.

[24] O. Gallo, R. Manduchi, A. Rafii, Robust curb and ramp detection for safe parking using the Canesta ToF camera, in: Proc. IEEE CVPR Workshops, Vol. 1-3, Anchorage, 2008, pp. 1558–1565. doi:10.1109/CVPRW.2008.4563165.

[25] S. B. Gokturk, A. Rafii, An occupant classification system eigen shapes or knowledge-based features, in: Proc. 19th IEEE Conf. Comput. Vis. Pattern Recognit., San Diego, 2005, pp. 57–57. doi:10.1109/CVPR.2005.410.

[26] F. Yuan, A. Swadzba, R. Philippsen, O. Engin, M. Hanheide, , S. Wachsmuth, Laser-based navigation enhanced with 3D time of flight data, in: Proc. IEEE Int. Conf. Robotics Autom., Kobe, 2009, pp. 2844–2850.

[27] K. D. Kuhnert, M. Stommel, Fusion of stereo-camera and PMD-camera data for real-time suited precise 3D environment reconstruction, in: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Vol. 1-12, Beijing, 2006, pp. 4780–4785. doi:10.1109/IROS.2006.282349.

[28] C. Netramai, M. Oleksandr, C. Joochim, H. Roth, Motion estimation of a mobile robot using different types of 3D sensors, in: Proc. 4th Int. Conf. Autonomic and Autonomous Syst., Gosier, 2008, pp. 148–153. doi:http://dx.doi.org/10.1109/ICAS.2008.39.

[29] B. Huhle, M. Magnusson, W. Strasser, A. J. Lilienthal, Registration of colored 3D point clouds with a kernel-based extension to the normal distributions transform, in: Proc. IEEE Int. Conf. Robotics Autom., Vol. 1-9, Pasadena, 2008, pp. 4025–4030. doi:10.1109/ROBOT.2008.4543829.

[30] A. Prusak, O. Melnychuk, H. Roth, I. Schiller, R. Koch, Pose estimation and map building with a time-of-flight camera for robot navigation, Int. J. Int. Syst. Tech. App. 5 (3-4) (2008) 355–364. doi:http://dx.doi.org/10.1504/IJISTA.2008.021298.

[31] A. Swadzba, A. Vollmer, M. Hanheide, S. Wachsmuth, Reducing noise and redundancy in registered range data for planar surface extraction, in: Proc. 19th IAPR Int. Conf. Pattern Recog., Vol. 1-6, Tampa, 2008, pp. 1219–1222.

[32] N. Vaskevicius, A. Birk, K. Pathak, J. Poppinga, Fast detection of polygons in 3D point clouds from noise-prone range sensors, in: Proc. IEEE Int. Workshop Safety, Security Rescue Robot., Rome, 2007, pp. 30–35. doi:10.1109/SSRR.2007.4381261.

[33] J. Poppinga, N. Vaskevicius, A. Birk, K. Pathak, Fast plane detection and polygonalization in noisy 3D range images, in: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Nice, 2008, pp. 3378–3383. doi:10.1109/IROS.2008.4650729.

[34] T. Prasad, K. Hartmann, W. Weihs, S. E. Ghobadi, A. Sluiter, First steps in enhancing 3D vision technique using 2D/3D sensors, in: Computer Vision Winter Workshop, Prague, 2006, pp. 82–86.

[35] B. Streckel, B. Bartczak, R. Koch, A. Kolb, Supporting structure from motion with a 3D-range-camera, in: Proc. 15th Scandinavian Conf. Imag. Anal., Vol. 4522, Aalborg, 2007, pp. 233–242.

[36] B. Huhle, S. Fleck, A. Schilling, Integrating 3D time-of-flight camera data and high resolution images for 3DTV applications, in: Proc. 1st IEEE Int. Conf. 3DTV, Kos Isl, 2007, pp. 289–292.

[37] S. E. Ghobadi, K. Hartmann, W. Weihs, C. Netramai, O. Loffeld, H. Roth, Detection and classification of moving objects-stereo or time-of-flight images, in: Proc. Int. Conf. Comput. Intell. Security, Vol. 1, Guangzhou, 2006, pp. 11–16. doi:10.1109/ICCIAS.2006.294082.

[38] S. Hussmann, T. Liepert, Robot vision system based on a 3D-ToF camera, in: Proc. 24th IEEE Instrum. Meas. Tech. Conf., Vol. 1-5, Warsaw, 2007, pp. 1405–1409. doi:10.1109/IMTC.2007.379356.

[39] S. A. Guomundsson, R. Larsen, B. K. Ersboll, Robust pose estimation using the Swissranger SR-3000 camera, in: Proc. 15th Scandinavian Conf. Imag. Anal., Vol. 4522, Aalborg, 2007, pp. 968–975.

[40] C. Beder, B. Bartczak, R. Koch, A comparison of PMD-cameras and stereo-vision for the task of surface reconstruction using patchlets, in: Proc. 21st IEEE Conf. Comput. Vis. Pattern Recognit., Vol. 1-8, Minneapolis, 2007, pp. 2692–2699. doi:10.1109/CVPR.2007.383348.

[41] T. Grundmann, Z. Xue, J. Kuehnle, R. Eidenberger, S. Ruehl, A. Verl, R. D. Zoellner, J. M. Zoellner, R. Dillmann, Integration of 6D object localization and obstacle detection for collision free robotic manipulation, in: Proc. IEEE/SICE Int. Sym. System Integration, Nagoya, 2008, pp. 66–71.

[42] U. Reiser, J. Kubacki, Using a 3D time-of-flight range camera for visual tracking, in: Proc. 6th IFAC/EURON Sym. Intell. Auton. Vehicles, Toulouse, 2007.

[43] S. Gächter, A. Harati, R. Siegwart, Incremental object part detection to-

ward object classification in a sequence of noisy range images, in: Proc. IEEE Int. Conf. Robotics Autom., Vol. 1-9, Pasadena, 2008, pp. 4037–4042. doi:10.1109/ROBOT.2008.4543831.
[44] J. Shin, S. Gachter, A. Harati, C. Pradalier, R. Siegwart, Object classification based on a geometric grammar with a range camera, in: Proc. IEEE Int. Conf. Robotics Autom., Kobe, 2009, pp. 2443–2448. doi:10.1109/ROBOT.2009.5152601.
[45] U. Klank, D. Pangeric, R. Rusu, M. Beetz, Real-time CAD model matching for mobile manipulation and grasping, in: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Saint Louis, 2009, pp. 290–296.
[46] Z. Marton, R. Rusu, D. Jain, U. Klank, M. Beetz, Probabilistic categorization of kitchen objects in table settings with a composite sensor, in: Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Saint Louis, 2009, pp. 4777–4784.
[47] T. Nakamura, K. Sugiura, T. Nagai, N. Iwahashi, T. Toda, H. Okada, T. Omori, Learning novel objects for extended mobile manipulation, J. of Int. and Rob. Sys. 66 (1-2) (2012) 187–204.
[48] J. Zhu, L. Wang, R. Yang, J. Davis, Fusion of time-of-flight depth and stereo for high accuracy depth maps, in: Proc. 22nd IEEE Conf. Comput. Vis. Pattern Recognit., Vol. 1-12, Anchorage, 2008, pp. 3262–3269. doi:10.1109/CVPR.2008.4587761.
[49] M. Lindner, M. Lambers, A. Kolb, Sub-pixel data fusion and edge-enhanced distance refinement for 2D/3D, Int. J. Int. Syst. Tech. App. 5 (3-4) (2008) 344–354. doi:10.1504/IJISTA.2008.021297.
[50] J. Fischer, B. Huhle, A. Schilling, Using time-of-flight range data for occlusion handling in augmented reality, in: Proc. Eurographics Sym. Virtual Environments, 2007, pp. 109–116.
[51] H. Surmann, A. Nüchter, J. Hertzberg, An autonomous mobile robot with a 3d laser range finder for 3d exploration and digitalization of indoor environments, Robotic Auton. Syst. 45 (3-4) (2003) 181–198. doi:10.1016/j.robot.2003.09.004.
[52] M. L. Hafiane, W. Wagner, Z. Dibi, O. Manck, Analysis and estimation of {NEP} and {DR} in {CMOS} tof-3d image sensor based on {MDSI}, Sensors and Actuators A: Physical 169 (1) (2011) 66 – 73. doi:10.1016/j.sna.2011.05.014.
[53] M. Lindner, A. Kolb, Calibration of the intensity-related distance error of the PMD ToF-camera, in: Proc. SPIE, Vol. 6764, Boston, 2007. doi:10.1117/12.752808.
[54] T. Kahlmann, F. Remondino, H. Ingensand, Calibration for increased accuracy of the range imaging camera Swissranger$^{TM}$, in: ISPRS Commission V Symposium, Dresden, 2006, pp. 136–141.
[55] G. Alenyà, B. Dellen, S. Foix, C. Torras, Robotized plant probing: Leaf segmentation utilizing time-of-flight data, Robotics Autom. Mag. 20 (3) (2013) 50–59.
[56] A. Ramisa, G. Alenyà, F. Moreno-Noguer, C. Torras, Determining where to grasp cloth using depth information, in: Proc. 14th Int. Conf. Cat. Assoc. Artificial Intelligence, Lleida, 2011.
[57] A. Ramisa, G. Alenyà, F. Moreno-Noguer, C. Torras, Using depth and appearance features for informed robot grasping of highly wrinkled clothes, in: Proc. IEEE Int. Conf. Robotics Autom., Saint Paul, 2012, pp. 1703–1708.

## Biographies

**Guillem Alenyà** is Research associate at the Spanish Scientific Research Council (CSIC). He received Ph.D. degree in 2007 from the Technical University of Catalonia (UPC) with a work on egomotion using active contours. Dr Alenyà was a Marie Curie fellow of the European Commission in the period 2002-2004 and has participated in the EU research projects PACO-PLUS, GARNICS, and IntellAct developing new robot perception, machine learning, and symbolic task planning algorithms.

**Sergi Foix** is PhD Student at Institut de Robòtica i Informàtica Industrial (CSIC-UPC) currently pursuing his PhD in Automatic Control, Robotics and Computer Vision at the Technical University of Catalonia (UPC). He received M.Sc. degrees in Intelligent Systems, and Automatic Control and Robotics from the University of Sunderland and the Technical University of Catalonia (UPC), respectively. His research interests include next-best-view planning under uncertainty, 3D object modelling and pose estimation for robot manipulation, and 3D feature extraction for point cloud registration.

**Carme Torras** is Research Professor at the Spanish Scientific Research Council (CSIC). She received M.Sc. degrees in Mathematics and Computer Science from the Universitat de Barcelona and the University of Massachusetts, respectively, and a Ph.D. degree in Computer Science from the Technical University of Catalonia (UPC). Prof. Torras has published five books and about two hundred papers in the areas of robotics, computer vision, geometric reasoning, and machine learning. She has been local project leader of several European projects, such as the 6th framework IP project "Perception, Action and COgnition through Learning of Object-Action Complexes" (PACO-PLUS), and the 7th framework STREP projects "GARdeNIng with a Cognitive System" (GARNICS) and "Intelligent observation and execution of Actions and manipulations" (IntellAct). She was awarded the Narcís Monturiol Medal of the Generalitat de Catalunya in 2000, and she became ECCAI Fellow in 2007 and member of Academia Europaea in 2010. Prof. Torras is Editor of the IEEE Transactions on Robotics.

**List of Figures**