ToF cameras for eye-in-hand robotics

G. Alenyà, S. Foix and C. Torras

I. 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. It has been used in a wide range of applications and here we will describe the lessons learned by using such cameras to perform robotic tasks, specifically in eye-in-hand configurations. In such configurations, the camera is attached to the end-effector of a robot manipulator, so that new images can be obtained by actively changing the point of view of the camera (Fig. 1).

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 manipulator robots 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 II 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.

Eye-in-hand configurations have been used extensively for object modelling [1], and more recently to enable robot interaction with the environment [2]. Section III presents and places in context some of the relevant works.

Regarding the quality of data, it is well known that raw ToF data is quite noisy and prone to several types of disturbances [3]. Some of them are systematic and can be calibrated, and others are non-systematic and sometimes can be filtered out. In Section IV systematic and non-systematic error sources are reviewed. Section V shows the combination of both ToF and color images to obtain colored point-clouds.

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

This work was supported by the Spanish Ministry of Science and Innovation under project PAU+ DPI2011-27510, by the EU Project IntellAct FP7-ICT2009-6-269959 and by the Catalan Research Commission through SGR-00155.

Some parts of this chapter were originally published in G. Alenyà , S. Foix and C. Torras. "ToF cameras for active vision in robotics," *Sensors and Actuators A: Physical*, 218: 10-22, 2014.

Authors are with Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain; {galenya,sfoix,torras}@iri.upc.edu

Fig. 1: Example of an eye-in-hand configuration, with a ToF camera attached to the manipulator end-effector, in this case a Chlorophyll meter.

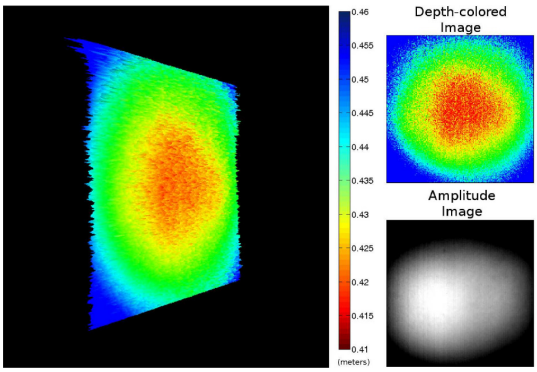


Fig. 2: Typical raw ToF image of a flat surface at short distance. Depth, ranging from 0.41 to 0.46 m, is color coded. Observe that over-illumination in the center leads to under-estimation of depth (shift to red), while under-illumination at borders causes over-estimation (shift to blue).

II. 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, which allows for the parallel measurement of its phase (cross-correlation), offset and amplitude [4]. Figure 2 shows a typical raw image of a flat surface with the depth values coded as different color values.

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

TABLE I: Specifications of different ToF cameras, and comparison with Kinect features.

The main characteristics of two ToF sensors, PMD CamCube 3 and Mesa Swissranger 4K are detailed in Table I. We include also the specifications of the Kinect sensors 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.

We focus this review on 3D perception for robotic manipulation and object modelling, thus resolution is an important factor. It is worth mentioning that 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 [5]. 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 autofocus 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. As will be presented in Section V, 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 [6]. 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.

III. 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 modelling [7], [8], precise surface reconstruction [9], and to grasp known [10] and unknown [11] objects. We 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.

TABLE II: ToF camera usage in scene-related tasks

Article	Topic	Advantages	Type of Sensor
Weingarten *et al.* [12]	Obstacle avoidance in static env.	3D at high rate	SR2 (depth)
May *et al.* [13], [14]	3D mapping	3D at high rate/No required Pan-Tilt	SR2 (depth)
May *et al.* [15]	Pose estimation/3D mapping	Registered depth-intensity	SR3 (depth + intensity)
Hedge and Ye [16]	Planar feature 3D mapping	3D at high rate/No required Pan-Tilt	SR3
Ohno *et al.* [17]	3D mapping	3D at high rate	SR2
Stipes *et al.* [18]	3D mapping / Point selection	Registered depth-intensity	SR3
May *et al.* [19]	3D mapping/SLAM	3D at high rate	SR3
Gemeiner *et al.* [20]	Corner filtering	Registered depth-intensity	SR3 (depth + intensity)
Thielemann *et al.* [21]	Navigation in pipelines	3D allow geometric primitives search	SR3
Sheh *et al.* [22]	Navigation in hard env.	3D at high rate	SR3 + inertial
Swadzba *et al.* [23]	3D mapping in dynamic env.	3D at high rate/Registered depth-intensity	SR3 (depth + intensity)
Acharya *et al.* [24] Gallo *et al.* [25]	Safe car parking	Improved depth range/3D at high rate	Canesta
Gortuk *et al.* [26]	Object classification (airbag app.)	light/texture/shadow independence	Canesta
Yuan *et al.* [27]	Navigation and obst. avoidance	Increased detection zone	SR3 + laser
Kuhnert and Stommel *et al.* [28]	3D reconstruction	Easy color registration	PMD + stereo
Netramai *et al.* [29]	Motion estimation	3D at high rate	PMD + stereo
Huhle *et al.* [30]	3D mapping	Easy registration of depth and color	PMD + color camera
Prusak et al. [31]	Obst. avoidance/Map building	Absolute scale/better pose estimation	PMD + spherical camera
Swadzba *et al.* [32]	3D mapping/Map optimisation	3D at high rate	SR3
Vaskevicius *et al.* [33] Poppinga [34]	Localization/Map optimisation	Neighbourhood relation of pixels No color restrictions	SR3

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 modelling, and to a lesser extent to enable object manipulation.

A table is provided in each section to summarise 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. Moreover, 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 applied to ToF depth images.

The depth-intensity image pair is also often used, exploiting the fact that both images are delivered already registered. In applications where the reduced resolution of a ToF camera is critical, it is complemented with other sensors, usually color cameras. ToF cameras are used in human environments because they are eye-safe and permit avoiding physical contact and dedicated markers or hardware.

Some of the reviewed works do not apply any calibration method to rectify the depth images. We believe that this explains several 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.

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.

A. 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 II). 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.* [12] 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.* [13], [14] 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 modelling 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 ex-*

tract geometric features. Most of the reviewed applications extract planar regions using both intensity and depth images. In [15], 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 [16] to detect badly conditioned plane detection, as horizontal planes in a staircase. Also Pathak *et al.* [35] 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.* [17] 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.* [18], 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.* [19] 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.* [20] 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.* [21] 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.* [22] have proposed a ToF based navigation system for a random stepfield terrain[2]. 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.* [23] 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 [24] Acharya *et al.* describe the system design of a ToF camera for backup obstacle detection. In [25] 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 [26]: adult, child, child seat or objects.

Fusion with other sensors. Some other 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.* [27] 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 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 [28] present a revision of their 3D environment reconstruction algorithm combining information from a stereo system and a ToF sensor. Later, Netramai *et al.* [29] 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.* [30] 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.* [31] 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 [36], [37], 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.* [32] 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

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

TABLE III: ToF camera usage in object-related tasks

Reference	Topic	Advantages	Type of Sensor
Ghobadi *et al.* [39]	Dynamic object detection and classification	Color and light independence	PMD
Hussmann and Liepert [40]	Object pose	Easy object/background segmentation	PMD
Guomundsson *et al.* [41]	Known object pose estimation	Light independent / Absolute scale	SR3
Beder *et al.* [42]	Surface reconstruction using patchlets	ToF easily combines with stereo	PMD
Fuchs and May [9]	Precise surface reconstruction	3D at high rate	SR3/O3D100 (Depth)
Dellen *et al.* [7] Foix *et al.* [8]	3D object reconstruction	3D at high rate	SR3 (Depth)
Kuehnle *et al.* [10]	Object recognition for grasping	3D allow geometric primitives search	SR3
Grundmann *et al.* [43]	Collision free object manipulation	3D at high rate	SR3 + stereo
Reiser and Kubacki [44]	Position based visual servoing	3D is simply obtained / No model needed	SR3 (Depth)
Gachter *et al.* [45] Shin *et al.* [46]	Object part detection for classification	3D at high rate	SR3 SR2
Marton *et al.* [47]	Object categorisation	ToF easily combines with stereo	SR4 + color
Saxena *et al.* [11]	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

Jacobs University [33], [34] 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 neighbourhood 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.* [38] 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.

B. 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 III

Comparison with stereovision. A classical solution in the area of object modelling is the use of calibrated stereo rigs. Therefore, initial works were devoted to their comparison with 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.* [39] 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 [40] also compared ToF and stereo vision for object pose computation. The key difference favourable 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.* [41]. 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.* [42] 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 modelling algorithms usually include a ToF camera mounted on the end-effector of a robotic arm. *Point cloud registration is more critical in object modelling than in scene modelling*. 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 [9] 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.* [7] 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. [8] 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 categorised 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.* [10] explore the use a ToF camera to recognise and locate 3D objects in the framework of the robotic manipulation system DESIRE. Objects are modelled with geometric primitives. Although they use depth images rectified up to some level, their system is not reliable enough. In a subsequent work [43] 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 [44] 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.* [45] propose to detect and classify objects by identifying their different parts. For example, chairs are modelled by finding their legs, which in turn are modelled 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.* [46] 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.

Depth information is very useful in cluttered environments to detect and grasp unknown objects: the 3D region of interest can be extracted easily, and 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.* [47] proposed a probabilistic categorisation 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.

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.* [11] 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 textureless 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 outlier 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 modelling 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.

IV. Depth measurement errors

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. Two of the most important systematic errors are *depth distorsion*, an offset that affects all the image and that is dependent on the measured depth (Fig. 3), and *built-in pixel errors*, which is a constant offset of each pixel independent of the measured depth. While systematic errors are compensated by calibration, non-systematic ones are minimised by filtering.

One of the know problems with ToF is 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 [3].

Our interest is to place the sensor very close to the scene components, usually in a range from 30 to 50cm. This high proximity makes ToF cameras easier to calibrate but more susceptible to some error sources. For example, *depth distortion* can be approximated linearly due to the reduced range, and the *built-in pixel errors* can be approximated with a look-up table. Special care should be taken to compensate errors due to saturation (amplitude-related) [8], light scattering [52] and multiple light reflections [9]. Note that newer ToF cameras allow to easily detect saturated pixels.

ToF cameras are evolving and a lot of work is being carried out to understand the source of errors and to compensate them. The next section presents a classification and short description of the different errors. A detailed ToF error description and classification can be found in [3].

A. Systematic Errors

Five types of systematic errors have been identified:

Depth distortion appears as a consequence of the fact that the emitted infra-red light cannot be generated in practice as theoretically planned (generally sinusoidal) due to irregularities in the modulation process. This type of error produces an offset that depends only on the measured depth for each pixel. Usually, the error plotted against the distance follows a sinusoidal shape[3] (see Fig. 3). This error is sometimes referred to as *wiggling* or *circular error*.

Built-in pixel-related errors arise from two main sources. On the one hand, errors due to different material properties in CMOS-gates. This produces a constant pixel-related distance offset, leading to different depths measured in two neighbour pixels corresponding to the same real depth. On the other hand, there are latency-related offset errors due to the capacitor charge time delay during the signal correlation process. This can be observed as a rotation on the whole scene (Fig. 4a) reporting wrong depth measurements. After calibration, the complete scene pose can be correctly recovered (Fig. 4b).

[3]This has been explained as due to perturbations in the measured signal phase caused by aliasing of odd harmonics contained in the emitted reference signal [53].

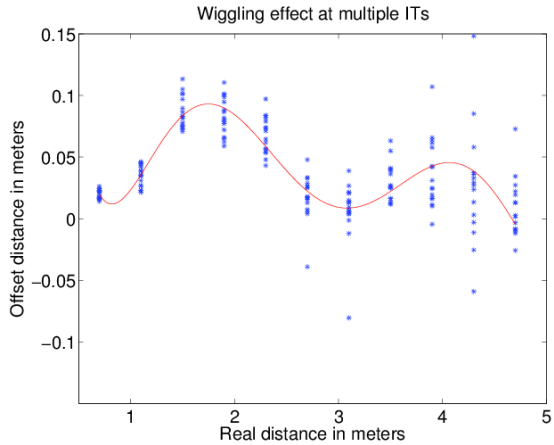


Fig. 3: Depth distortion offset (wiggling effect). (Blue) Measurements captured with a SR3100 ToF camera at several integration times (2ms - 32 ms). (Red) 6-degree polynomial approximated function.

Amplitude-related errors occur due to low or saturated reflected amplitudes. Low amplitude appears more often in the border of the image as the emitted light is lower that in the center where the depth is overestimated. Contrarily, when the object is too close to the camera, saturation can appear and depth is underestimated (observe Fig. 2). Moreover, amplitude-related errors occur due to differences in the object reflectivity, causing differences in the amount of reflected light, and thus yielding different depth measurements for the same constant distance. This effect can be recognized in Fig. 4c. The image corresponds to a typical calibration pattern: a checker-board of white and black squares. Observe the difference in depth of the points corresponding to squares of each color.

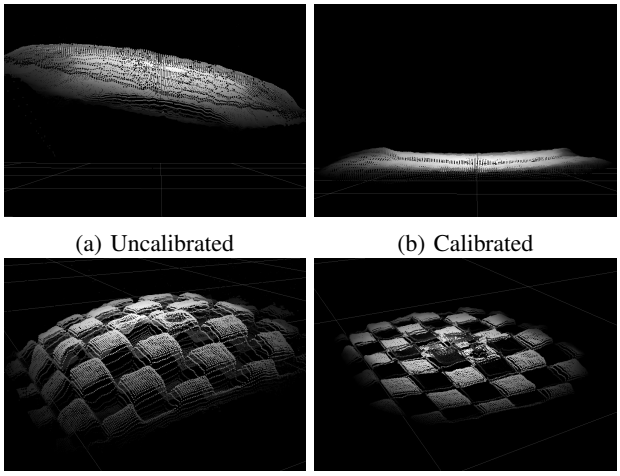
Both pixel-related errors (depth and amplitude) cause a constant depth miss-measurements and can be compensated by means of pixel-offset-based calibration methods as the so-called Fixed Pattern Noise (FPN) [54]. After the correction determined by calibration, the checkerboard 3D structure can be recovered (Fig. 4d).

Integration time-related error. Integration time (IT) can be selected by the user. It has been observed that for the same scene different IT cause different depth values in the entire scene. The main reason for this effect is still subject of investigation.

Temperature-related errors happen because internal camera temperature affects depth processing, explaining why some cameras include an internal fan. Depth values suffer from a drift in the whole image until the temperature of the camera is stabilised.

B. Non-systematic Errors

Four non-systematic errors can also be identified in depth measurements with ToF cameras, the occurrence of the last three being unpredictable.



(a) Uncalibrated (b) Calibrated

(c) Uncalibrated (d) Calibrated

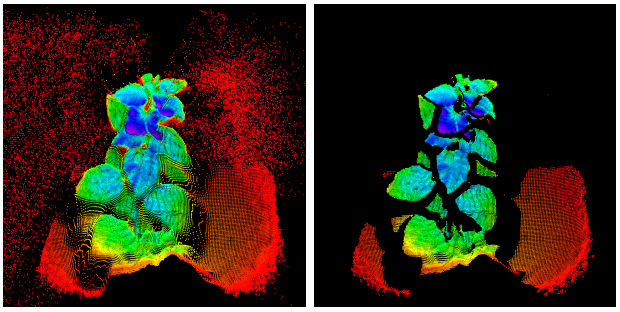
Fig. 4: 3D image of a planar surface and a white/black checkerboard. (a) surface should be horizontal, but *built-in pixel-related* error causes a distortion. (b) Once calibrated, the orientation of the plane and the depth of individual points is corrected. (c) Observe the difference in depth between the squares of each color. (d) The calibrated image is rectified taking into account built-in and amplitude errors.

Signal-to-noise ratio distortion appears in scenes not uniformly illuminated. Low illuminated areas are more susceptible to noise than high illuminated ones. This type of error is highly dependent on the amplitude, the IT parametrisation and the depth uniformity of the scene. Non-uniform depth over the scene can lead to low-amplitude areas (far objects) that will be highly affected by noise. Usually the IT is calculated to optimally increase accuracy on the distance range of the working scene area.

Multiple light reception errors appear due to the interference of multiple light reflections captured at each sensor's pixel. These multiple light reflections depend on the geometric shape of the objects in the scene and can have two origins. The more obvious is due to concavities that cause multiple reflections. The other one is produced when different depths project to the same pixel, it is more obvious in the edges of the objects, and generates the so called flying points between foreground and background. Flying points can be detected and filtered out (see Fig 5).

Light scattering effect arises due to multiple light reflections between the camera lens and its sensor. This effect produces a depth underestimation over the affected pixels, because of the energy gain produced by its neighbouring pixel reflections [55]. Errors due to light scattering are only relevant when nearby objects are present in the scene. The closer an object, the higher the interference [56]. This kind of errors are hard to rectify, but some ToF cameras permit to identify over-exposed pixels using some control flags.

Motion blurring, present when traditional cameras are used in dynamic environments, appears also with ToF cameras. This is due to the physical motion of the objects or the camera during the integration time used for sampling.



(a) Raw image (b) Filtered

Fig. 5: Reduction of noise by filtering pixels using a flying-points detector and depth threshold filtering.

V. Coloring depth points

The combination of ToF images and color images can be performed to obtain colored point-clouds [57], like the ones delivered by Kinect, using the extrinsic calibration parameters between both cameras.

Traditionally extrinsic calibration has been addressed by considering the intensity image of the ToF camera and using classical stereo calibration algorithms [58], [50], [59], [43]. However, the characteristic low resolution of this type of camera leads to a poor localization of the calibration pattern points and the obtained calibration parameters are usually noisy.

Thus, the idea is to take advantage of depth information when calibrating. Once a color camera has been calibrated with a known pattern, reconstruction of the calibration poses is possible, and this can be used to find better extrinsic parameters [60]. A software to calibrate one or multiple color cameras with a ToF camera using this principle is available [61]. This algorithm also includes a depth calibration model that represents the depth deviation as a polynomial function, similar to [9].

Once the extrinsic parameters of the coordinate transformation between a color camera and a ToF camera have been obtained, data fusion is possible. The easy part is to find the correspondences between them and put color to the depth image, but more can be done. Due to the difference in resolution (i.e., 204×204 pixels a CamCube image, and 640×480 a color image), between each pair of neighbour points in the ToF image there are several points in the color image. As a consequence, these points can be interpolated to obtain a dense depth map [58] where all the color points can be used.

Bartczak *et al.* [62] use a 3D surface mesh that is rendered into the color camera view as an alternative algorithm to obtain a dense depth map. Huhle *et al.* [38] present a completely different approach, where the dense depth map is obtained using a Markov Random Field (MRF). Depths are represented in the model taking explicitly into account the discontinuities, which are used as a prior to perform the alignment.

As it is typical in stereovision, some scene points are seen by one camera but not by the other due to their slightly different viewpoints. Consequently, for some depth points it is

Fig. 6: Calibration errors produce bad colored points at the edge of the leaf. Additionally, observe the wrong color assignment in some background points, marked automatically in red, corresponding to the shadow of the leaf. This problem arises as the optical axes of the depth camera and the color camera are not the same and some depth points have no correspondence in the color image. Other sensor combinations, like Kinect, suffer the same problem.

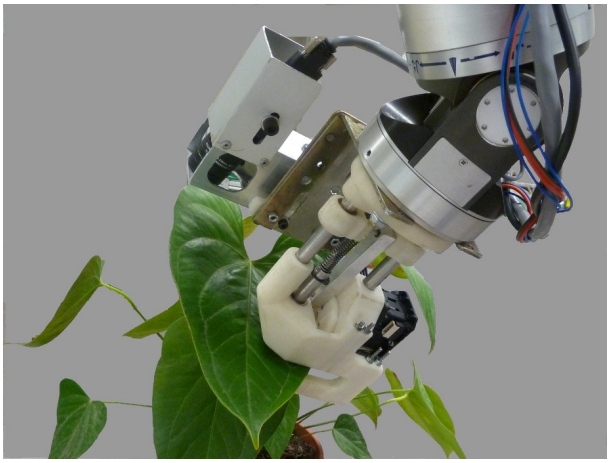
impossible to find their corresponding one in the color image. Note that these occlusions appear mainly for close objects, precisely our scenario.

Figure 6 is a detail of an image acquired with a CamCube + color camera sensor (Fig 7a). In this example, occluded points are detected and colored in red using a Z_buffer approach. First, the point cloud is transformed to the RGB camera reference frame using the extrinsic transformation matrix $\mathbf{F}$. Ideally, this leads to 3D points projecting to the corresponding pixel in the color image. In the case of occlusion, only the point that is closer to the camera is stored in the Z_buffer. However, as the ToF camera has a lower resolution than the color camera, it is possible that two 3D points (namely, the foreground and the occluded background points) do not project exactly onto the same color point, so no one is removed. This can lead to a mosaic of foreground-background pixels in the regions where occlusions occur. A neighbourhood region can be taken into account to build the Z_buffer, so that the depth of neighbours determines whether occlusions are to be considered.

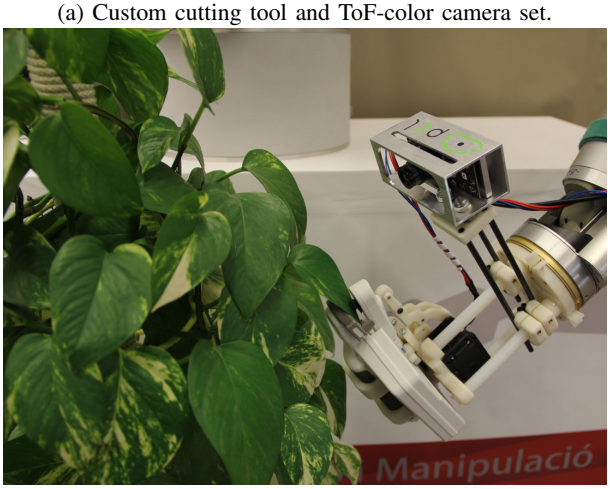
To completely avoid the occlusion problem, the ToF and the color camera optical axes should be the same. This can be accomplished using a beam splitter between the two cameras mounted at 90° [63], [64].

VI. Applications

Some examples of the applicability of ToF cameras in eye-in-hand configuration are presented in this section. Three 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), the ability to acquire accurate views of particular details of the scene, and the ability to disambiguate scenes.



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

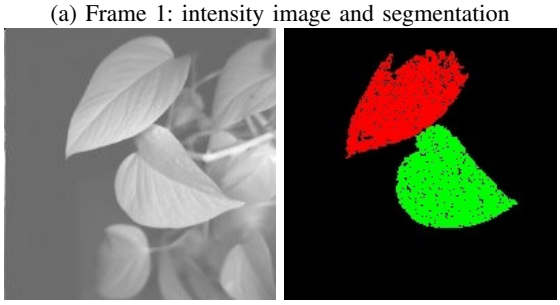


(b) Chlorophyll meter and ToF camera.

Fig. 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.

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 to transport plants to a measuring cabin, where a set of sensors perform 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

(a) Frame 1: intensity image and segmentation



(b) Frame 2: intensity image and segmentation

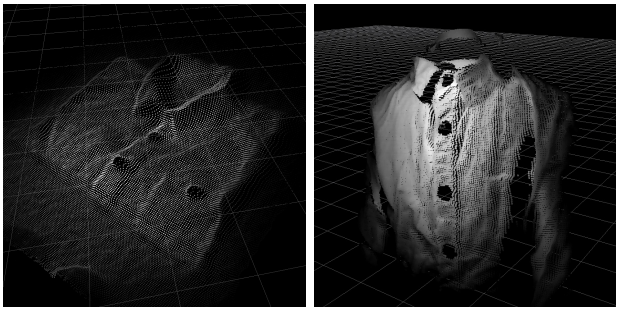
Fig. 8: Scene containing a detected leaf occlusion. After changing the point of view the occluded leaf is discovered and more characteristics (e.g., leaf area) can be measured. Depending on the particular leaf arrangement, it is not always possible to completely observe the occluded leaf.

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).

a) 3D structure: One of the objectives in plant phenotyping is to gather as much information as possible about each specimen, preferably 3D relevant information to enable its 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. 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.

Figure 8 shows an example of two leaves, the one in the foreground partially occluding the one in the background. Segmentation using only depth information is straightforward. Observe that the background leaf can be better observed after a camera motion. 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.

b) 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



(a) Folded (b) On a mannequin

Fig. 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.

(a) RGB color image (b) Colored depth image

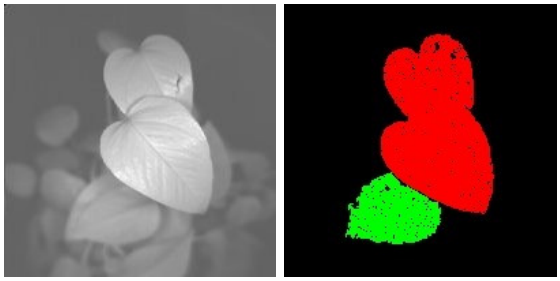
Fig. 10: Detail of a plant. Observe that the stems, even if they are thin, are correctly acquired.

relevant details can be better focused.

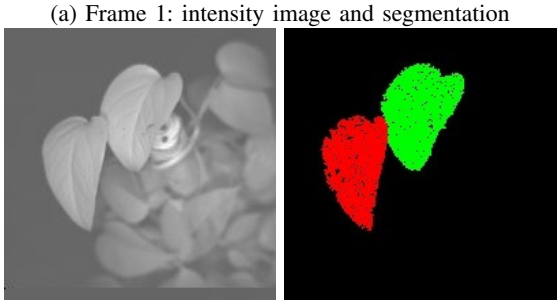
Figure 9 shows the image of a shirt in two different configurations: folded and hanged. Here the task is to grasp the shirt from the collar to allow the robot to hang the shirt in a hanger. Observe that in both configurations the details of the collar, the buttons and small winkles are visible. In the hanged shirt the sleeves are identifiable as well. Previous works have shown that this 3D structure can be used to identify wrinkles [65] and also the collar structure, using computer vision algorithms [66].

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.

c) Disambiguation: Segmentation algorithms use different parameters to adapt 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 11 shows an example, where in the first view the segmentation algorithm,



(a) Frame 1: intensity image and segmentation



(b) Frame 2: intensity image and segmentation

Fig. 11: Scene containing a possible merging of leaves. After changing the point of view, the ambiguity is clarified and two leaves are detected instead of one. Depending on the particular leaf arrangement, it is not always possible to completely disambiguate the occluded leaf.

that uses depth similarity between adjusted surfaces, fails to distinguish two different leaves. Using a next-best-view algorithm [8], 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.

VII. 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 minimised 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. We have highlighted here eye-in-hand configurations, where the sensor is mounted on the end-effector of a robot manipulator and it is placed at a short-distance from the target object. We have provided experimental evidence of the effectivity of such approach in three tasks: 3D structure recovering of plants, acquisition of detailed views, and disambiguation.

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 show great potential in the development of data feature extraction, registration, reconstruction, planning and optimisation 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 mathematical 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 strong noise 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 position 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.

Other concerns include ambient light noise, motion artifacts and high-reflectivity surfaces in the scene. Ambient light may contain unwanted light of the same wavelength as that of the ToF light source, which may cause false sensor measurements. Frequency-based filters can be used in order to minimise this effect. Motion artifacts are errors caused by receiving light from different depths at the same time due to object motion in the scene. This type of errors are mostly observed around the edges of the moving object and can be attenuated either by increasing the frame rate, or by correction using motion estimation. Finally, errors due to the coexistence of low-reflective and high-reflective objects (mirroring effect) can be addressed by combining multiple exposure settings.

References

[1] S. Foix, S. Kriegel, S. Fuchs, G. Alenyà, and C. Torras, "Information-gain view planning for free-form object reconstruction with a 3D ToF camera," in *Internatial Conference on Advanced Concepts for Intelligent Vision Systems*, Sep. 2012.

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

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

[4] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE J. Quantum Electron.*, vol. 37, no. 3, pp. 390–397, Mar. 2001.

[5] "PMD-Cameras, http://www.pmdtec.com." PMDTechnologies GmbH, 2009.

[6] S. F. W. Kazmi and G. Alenyà, "Plant leaf imaging using time of flight camera under sunlight, shadow and room conditions," in *IEEE International Symposium on Robotic and Sensors Environments*, 2012, pp. 192–197.

[7] B. Dellen, G. Alenyà, S. Foix, and 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, Feb. 2009, pp. 368–372.

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

[9] S. Fuchs and S. May, "Calibration and registration for precise surface reconstruction with time of flight cameras," *Int. J. Int. Syst. Tech. App.*, vol. 5, no. 3-4, pp. 274–284, 2008.

[10] J. U. Kuehnle, Z. Xue, M. Stotz, J. M. Zoellner, A. Verl, and R. Dillmann, "Grasping in depth maps of time-of-flight cameras," in *Proc. Int. Workshop Robot. Sensors Environ.*, Otawa, Oct. 2008, pp. 132–137.

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

[12] J. Weingarten, G. Gruener, and R. Siegwart, "A state-of-the-art 3D sensor for robot navigation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, vol. 3, Sendei, Sep. 2004, pp. 2155–2160.

[13] S. May, B. Werner, H. Surmann, and K. Pervolz, "3D time-of-flight cameras for mobile robotics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Beijing, Oct. 2006, pp. 790–795.

[14] S. May, K. Pervolz, and H. Surmann, *Vision systems: applications*. I-Tech Education and Publishing, 2007, ch. 3D cameras: 3D computer vision of wide scope.

[15] S. May, D. Droeschel, D. Holz, C. Wiesen, and S. Fuchs, "3D pose estimation and mapping with time-of-flight cameras," in *Proc. IEEE/RSJ IROS Workshop on 3D-Mapp.*, Nice, Sep. 2008.

[16] G. Hedge and 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, Oct. 2009, pp. 4034–4039.

[17] K. Ohno, T. Nomura, and S. Tadokoro, "Real-time robot trajectory estimation and 3D map construction using 3D camera," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Beijing, Oct. 2006, pp. 5279–5285.

[18] J. Stipes, J. Cole, and J. Humphreys, "4D scan registration with the SR-3000 LIDAR," in *Proc. IEEE Int. Conf. Robotics Autom.*, Pasadena, May 2008, pp. 2988–2993.

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

[20] P. Gemeiner, P. Jojic, and 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, Oct. 2009, pp. 5711–5716.

[21] J. Thielemann, G. Breivik, and A. Berge, "Pipeline landmark detection for autonomous robot navigation using time-of-flight imagery," in *Proc. IEEE CVPR Workshops*, vol. 1-3, Anchorage, Jun. 2008, pp. 1572–1578.

[22] R. Sheh, M. W. Kadous, C. Sammut, and B. Hengst, "Extracting terrain features from range images for autonomous random stepfield traversal," in *Proc. IEEE Int. Workshop Safety, Security Rescue Robot.*, Rome, Sep. 2007, pp. 24–29.

[23] A. Swadzba, N. Beuter, J. Schmidt, and G. Sagerer, "Tracking objects in 6D for reconstructing static scenes," in *Proc. IEEE CVPR Workshops*, vol. 1-3, Anchorage, June 2008, pp. 1492–1498.

[24] S. Acharya, C. Tracey, and 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, Jun. 2008, pp. 1552–1557.

[25] O. Gallo, R. Manduchi, and A. Rafii, "Robust curb and ramp detection for safe parking using the Canesta ToF camera," in *Proc. IEEE CVPR Workshops*, vol. 1-3, Anchorage, June 2008, pp. 1558–1565.

[26] S. B. Gokturk and A. Rafii, "An occupant classification system eigen shapes or knowledge-based features," in *Proc. 19th IEEE Conf. Comput. Vis. Pattern Recognit.*, San Diego, June 2005, pp. 57–57.

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

[28] K. D. Kuhnert and 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, Oct. 2006, pp. 4780–4785.

[29] C. Netramai, M. Oleksandr, C. Joochim, and H. Roth, "Motion estimation of a mobile robot using different types of 3D sensors," in *Proc. 4th Int. Conf. Autonomic and Autonomous Syst.*, Gosier, Mar. 2008, pp. 148–153.

[30] B. Huhle, M. Magnusson, W. Strasser, and 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, May 2008, pp. 4025–4030.

[31] A. Prusak, O. Melnychuk, H. Roth, I. Schiller, and R. Koch, "Pose estimation and map building with a time-of-flight camera for robot navigation," *Int. J. Int. Syst. Tech. App.*, vol. 5, no. 3-4, pp. 355–364, 2008.

[32] A. Swadzba, A. Vollmer, M. Hanheide, and 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, Dec. 2008, pp. 1219–1222.

[33] N. Vaskevicius, A. Birk, K. Pathak, and 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, Sep. 2007, pp. 30–35.

[34] J. Poppinga, N. Vaskevicius, A. Birk, and K. Pathak, "Fast plane detection and polygonalization in noisy 3D range images," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nice, Sep. 2008, pp. 3378–3383.

[35] K. Pathak, N. Vaskevicius, J. Poppinga, S. Schwertfeger, M. Pfingsthorn, and 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, Oct. 2009, pp. 1150–1155.

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

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

[38] B. Huhle, S. Fleck, and 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, May 2007, pp. 289–292.

[39] S. E. Ghobadi, K. Hartmann, W. Weihs, C. Netramai, O. Loffeld, and H. Roth, "Detection and classification of moving objects-stereo or time-of-flight images," in *Proc. Int. Conf. Comput. Intell. Security*, vol. 1, Guangzhou, Nov. 2006, pp. 11–16.

[40] S. Hussmann and T. Liepert, "Robot vision system based on a 3D-ToF camera," in *Proc. 24th IEEE Instrum. Meas. Tech. Conf.*, vol. 1-5, Warsaw, May 2007, pp. 1405–1409.

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

[42] C. Beder, B. Bartczak, and 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, June 2007, pp. 2692–2699.

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

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

[45] S. Gächter, A. Harati, and R. Siegwart, "Incremental object part detection toward object classification in a sequence of noisy range images," in *Proc. IEEE Int. Conf. Robotics Autom.*, vol. 1-9, Pasadena, May 2008, pp. 4037–4042.

[46] J. Shin, S. Gachter, A. Harati, C. Pradalier, and R. Siegwart, "Object classification based on a geometric grammar with a range camera," in *Proc. IEEE Int. Conf. Robotics Autom.*, Kobe, May 2009, pp. 2443–2448.

[47] Z. Marton, R. Rusu, D. Jain, U. Klank, and 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, Oct. 2009, pp. 4777–4784.

[48] J. Zhu, L. Wang, R. Yang, and 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, June 2008, pp. 3262–3269.

[49] M. Lindner, M. Lambers, and A. Kolb, "Sub-pixel data fusion and edge-enhanced distance refinement for 2D/3D," *Int. J. Int. Syst. Tech. App.*, vol. 5, no. 3-4, pp. 344–354, 2008.

[50] J. Fischer, B. Huhle, and A. Schilling, "Using time-of-flight range data for occlusion handling in augmented reality," in *Proc. Eurographics Sym. Virtual Environments*, Sep. 2007, pp. 109–116.

[51] H. Surmann, A. Nüchter, and J. Hertzberg, "An autonomous mobile robot with a 3d laser range finder for 3d exploration and digitalization of indoor environments," *Robotic Auton. Syst.*, vol. 45, no. 3-4, pp. 181–198, 2003.

[52] J. Mure-Dubois and H. Hügli, "Real-time scattering compensation for time-of-flight camera," in *Proc. 5th Int. Conf. Comput. Vis. Syst.*, Bielefeld, March 2007.

[53] R. Lange, "3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technoloty," Ph.D. dissertation, Univ. Siegen, Germany, 2000.

[54] T. Kahlmann, F. Remondino, and H. Ingensand, "Calibration for increased accuracy of the range imaging camera Swissranger^{TM}," in *ISPRS Commission V Symposium*, Dresden, Sep. 2006, pp. 136–141.

[55] T. Kahlmann and H. Ingensand, "Calibration and development for increased accuracy of 3D range imaging cameras," *J. Appl. Geodesy*, vol. 2, no. 1, pp. 1–11, 2008.

[56] W. Karel, "Integrated range camera calibration using image sequences from hand-held operation," in *Proc. ISPRS Conf.*, vol. 37, Beijing, Jul. 2008, pp. 945–952.

[57] 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. Robotics Autom.*, Shanghai, May 2011, pp. 3408–3414.

[58] M. Lindner, A. Kolb, and K. Hartmann, "Data-fusion of PMD-based distance-information and high-resolution RGB-images," in *Proc. Int. Sym. Signals Circuits Syst.*, vol. 1-2, Lasi, July 2007, pp. 121–124.

[59] S. A. Guomundsson, H. Aanæs, and R. Larsen, "Fusion of stereo vision and time-of-flight imaging for improved 3D estimation," *Int. J. Int. Syst. Tech. App.*, vol. 5, no. 3-4, pp. 425–433, 2008.

[60] I. Schiller, C. Beder, and R. Koch, "Calibration of a PMD camera using a planar calibration object together with a multi-camera setup," in *Proc. ISPRS Conf.*, vol. 37. Part B3a, Beijing, Jul. 2008, pp. 297–302.

[61] "http://mip.informatik.uni-kiel.de," 2009.

[62] B. Bartczak, I. Schiller, C. Beder, and R. Koch, "Integration of a time-of-flight camera into a mixed reality system for handling dynamic scenes, moving viewpoints and occlusions in real-time," in *Proc. Int. Sym. 3D Data Processing, Visualization and Transmission (3DPVT)*, Atlanta, Jun. 2008.

[63] S. Ghobadi, O. Loepprich, F. Ahmadov, and J. Bernshausen, "Real time hand based robot control using multimodal images," *Int. J. Comput. Science*, vol. 35, no. 4, 2008.

[64] J.-H. Cho, S.-Y. Kim, Y.-S. Ho, and K. Lee, "Dynamic 3D human actor generation method using a time-of-flight depth camera," *IEEE Trans. Consumer Electron.*, vol. 54, no. 4, pp. 1514–1521, Nov. 2008.

[65] A. Ramisa, G. Alenyà, F. Moreno-Noguer, and C. Torras, "Determining where to grasp cloth using depth information," in *Proc. 14th Int. Conf. Cat. Assoc. Artificial Intelligence*, Lleida, Oct. 2011.

[66] A. Ramisa, G. Alenyà, F. Moreno-Noguer, and C. Torras, "Using depth and appearance features for informed robot grasping of highly wrinkled clothes," in *Proc. IEEE Int. Conf. Robotics Autom.*, Saint Paul, May 2012, pp. 1703–1708.