Noname manuscript No.
(will be inserted by the editor)

Using ToF and RGBD cameras for 3D robot perception and manipulation in human environments

G. Alenyà · S. Foix · C. Torras

Received: date / Accepted: date

Abstract Robots, traditionally confined into factories, are nowadays moving to domestic and assistive environments, where they need to deal with complex object shapes, deformable materials, and pose uncertainties at human pace. To attain quick 3D perception, new cameras delivering registered depth and intensity images at a high frame rate hold a lot of promise, and therefore many robotics researchers are now experimenting with structured-light RGBD and Time-of-Flight (ToF) cameras.

In this paper both technologies are critically compared to help researchers to evaluate their use in real robots. The focus is on 3D perception at close distances for different types of objects that may be handled by a robot in a human environment. We review three robotics applications. The analysis of several performance aspects indicates the complementarity of the two camera types, since the user-friendliness and higher resolution of RGBD cameras is counterbalanced by the capability of ToF cameras to operate outdoors and perceive details.

Keywords manipulation in human environments; deformable objects; complex object shapes; close distance perception; Time-of-Flight cameras

This research is partially funded by the EU GARNICS project FP7-247947, by CSIC project MANIPlus 201350E102, by the Spanish Ministry of Science and Innovation under project PAU+ DPI2011-27510, and the Catalan Research Commission under grant SGR-155.

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

1 Introduction

The technology of 3D cameras has quickly evolved in recent years, yielding off-the-shelf devices with great potential in many scientific fields ranging from virtual reality to surveillance and security. Within robotics, these cameras open up the possibility of real-time robot interaction in human environments, by offering an alternative to time-costly procedures such as stereovision and laser scanning. Time-of-Flight (ToF) cameras, provided by Mesa Imaging and PMD Technologies among others, appeared first and attracted a lot of attention with dedicated workshops (e.g., within CVPR'08) and a quickly growing number of papers at major conferences. Recently, RGBD cameras, composed by a classical RGB camera and a depth sensor using the Light Coding technology provided by PrimeSense and based on Structured Light (SL), have received even greater attention, because of their low cost and simplicity of use. A proof of this is the organization of several RGBD workshops within the main robotics conferences (e.g. RSS'10,'11,'12, ICRA'12 and IROS'12), and the numerous special issues that are currently being edited at renowned journals.

ToF cameras started to be used in robotics as a commercially available sensor around 2004. Their main use was in robot navigation and other long range tasks, in short range tasks like object modeling and grasping, and in less extend in human activity recognition and robot-human interaction (for a complete survey see [3, 6]). The appearance of RGBD cameras around 2011 revolutionized robotics applications, as the sensor is very easy to use and offers data of enough quality for most applications. RGBD has become widely used in human activity recognition, but also in mobile robotics and object recognition. For a compendium of the last

developments see [7]. Another proof of its acceptance is the appearance of numerous databases containing images of a large number of objects that have been made publicly available [10]. Meanwhile, ToF cameras technology has continued to evolve, and due to their characteristics, they still can be interesting for some robotics applications. In particular, we show their usage in close range tasks like those typical in robot manipulation and eye-in-hand robotics.

The evaluation of 3D cameras presented herein was triggered by three projects entailing manipulation of objects in three different environments: kitchen, botanics, and textile. Within the former European project PACO-PLUS, we studied the use of ToF cameras to assist robot learning of manipulation skills in a kitchen environment. Since this entailed mobile manipulation of rigid objects guided by a human teacher, we surveyed near one hundred previous works in three scenarios of application, namely scene-related tasks involving mobile robots in large environments, object-related tasks entailing robot interaction at short distances, and human-related tasks dealing with face, hand and body recognition for robot-human interfaces. Our conclusion was that ToF cameras, despite their relatively low resolution, seem especially adequate for mobile robotics and real-time applications in general, and in particular for the automatic acquisition of 3D models requiring sensor motion and on-line involved computations, which was the target application finally developed [5].

The European project GARNICS aimed to automatically monitor large botanic experiments to determine the best treatments (watering, nutrients, sunlight) to optimize predefined aspects (growth, seedling, flowers) and to eventually guide robots, like the one in Fig. 1, to interact with plants in order to obtain samples from leaves to be analyzed or even to perform some pruning. Here the interest was focused on 3D model acquisition of deformable objects (leaves) and their subsequent manipulation, i.e., the second scenario above.

Color vision is helpful to extract some relevant plant 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 or ambiguous, thus planning the next best 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 30 frames-per-second) a good option in front of other depth measuring procedures, such as stereovision or laser scanners.

The ongoing project PAU+ tackles the problem of modeling and manipulation of deformable objects, like textiles. This is a challenging task, since capturing the state of highly deformable objects, or detecting specific parts of such objects, is complex. Depth information plays an important role in this context, as it enables the development of 3D spatial descriptors that can be combined with existing appearance descriptors. Our conclusion is that the ability of combining color and depth is crucial and consequently RGBD cameras are a good option to improve the detection of specific parts, as well as to characterize the state of a textile e.g. building a map of the actual wrinkles [18].

In this paper we undertake a comparative assessment of the usefulness of both ToF and RGBD cameras to acquire (possibly deformable) object models at close distances for robot manipulation tasks. The main objective is to present in a comprehensive way *practical* aspects of both technologies, and to evaluate not just physical sensor features (e.g,, field of view, delivered image size, frame rate, focus and integration time), but also experimental performance aspects, such as operational distance range, calibration requirements, precision, occlusions, illumination conditions and ease of use, among others.

We contribute the practical learned lessons and the conclusions derived from our experience with 3D cameras in the three aforementioned projects; specifically, in next-best-view planning for object modeling and for best interaction with plants, and in perception for textile manipulation.

2 3D cameras evaluated

3D images are commonly represented as images with color codifying the depth, or as projections of 3D pointclouds. Figure 2 shows typical 3D images of a plant leaf acquired with both types of sensors. The main characteristics of two ToF cameras, PMD CamCube 3 and Mesa Swissranger 4K, as well as the most common RGBD cameras (Kinect, Asus Xtion, Carmine) are detailed in Table 1. Classically, 3D was obtained with a passive stereo system, and it is known to be still a very good alternative when viewing textured objects. The market offers already calibrated stereo systems ready to be used off-the-shelf. These systems are also RGBD sensors, as the correspondence method used to determine the depth for each pixel provide also the color component. Abusing of terminology, in this work RGBD will denote only Kinect-like cameras. We refer to [12] for a detailed review of stereo vision algorithms compared to ToF cameras in the context of plant-leaf segmentation.

(a) WAM robot arm with light PMD Camboard and chlorophyll sensor.

(b) Detailed view of PMD Camcube camera and probing tool.

Fig. 1 Experimental setup for the GARNICS project with a robot holding a ToF camera and two tools for (a) chlorophyll measurement and (b) leaf sample cutting, respectively.

A ToF camera simultaneously delivers intensity and range values for every pixel. Depth measurements are based on the well-known time-of-flight principle. A radio-frequency modulated-light field is emitted by the system and then reflected back to the sensor, which permits measuring in parallel its phase (cross-correlation), offset and amplitude [14].

Kinect uses an infrared structured light emitter to project a pattern into the scene and a camera to acquire the image of that pattern; then depth is computed by means of structured light algorithms. Additionally, Kinect integrates a high resolution color camera.

Kinect was developed with the idea of robust interactive human body tracking and large efforts have been made in this direction [19]. The community rapidly started to use Kinect after its protocol was unofficially made available, first with the same idea of human interaction and afterwards in other areas, like robot navigation (see the TurtleBot robot) and scene modeling (see Faro Scenect 5.2 software or the free implementation of the KinFu algorithm). Later, the official library was made public through the OpenNi organization and now it is integrated in the major perception libraries, like OpenCV [1] and PCL[2].

PCL has become a standard in 3D vision. Both RGBD and ToF cameras produce data in a format that permits taking advantage of the methods implemented in this library. PCL also offers several procedures for data storage, visualization and analysis, which have been useful in the aforementioned projects. It is worth mentioning that ToF data is noisier and, consequently, PCL-filtering modules have been helpful.

[1] http://opencv.org/

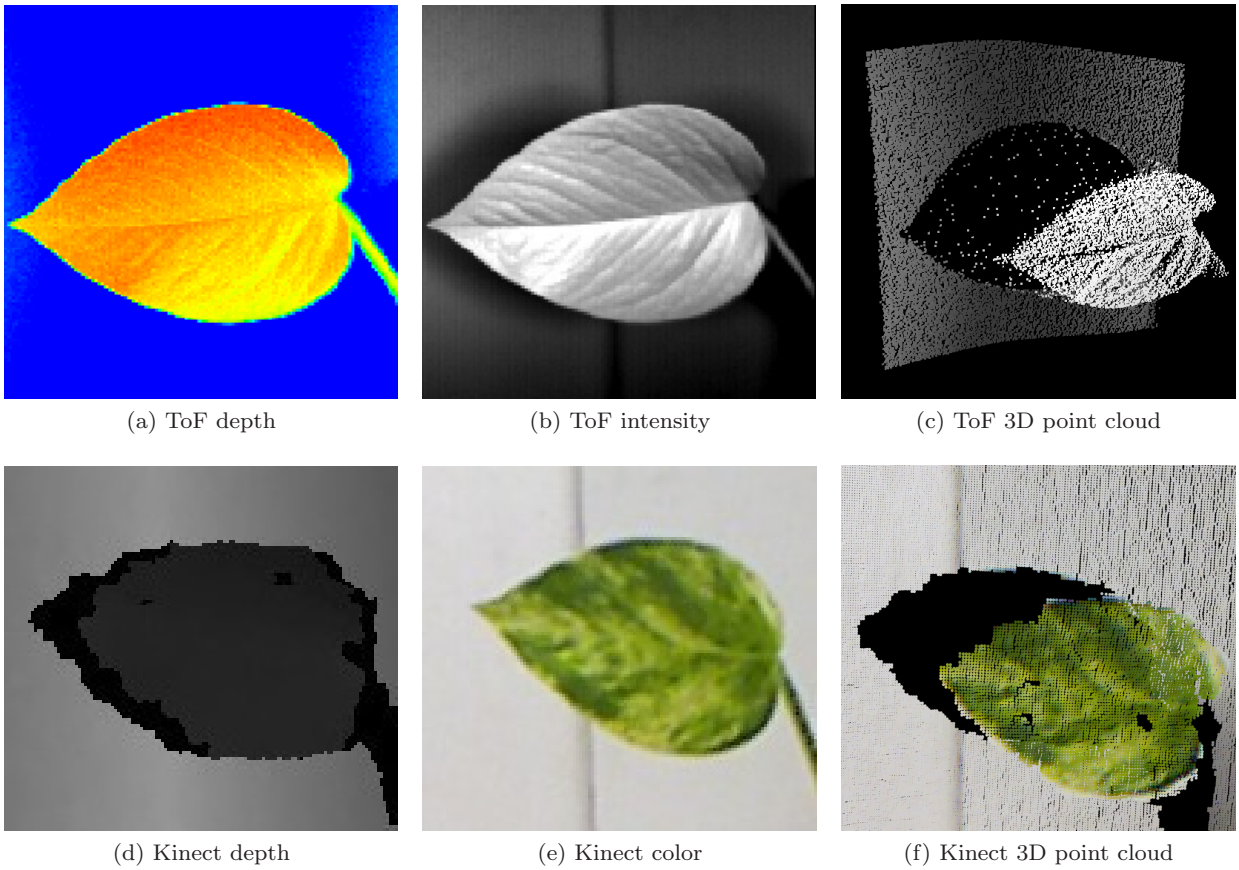
[2] http://pointclouds.org/

*Frame rate and resolution.*All cameras can deliver depth images at reasonably high frame rates. Their main difference is in depth image resolution: ToF's is typically around 200 x 200 (40000 depth points), while RGBD is 640 x 480 (307200 points). A new RGBD camera is expected to appear with a resolution of 1240x980. The functioning principle of RGBD cameras relies on the projection of a pattern of spots onto the scene (patent "Depth mapping using projected patterns" - 20100118123). Naturally, depth measurements can be performed only at the sensed spots, so the real resolution is restricted to the number of such spots. The actual figures are unknown, but it is accepted that approximately one out of every 9 pixels in the image is bright, leading to a real resolution of approximately 34650 pixels. Depth for the remaining pixels is interpolated up to VGA resolution.

Working distance We focus this work on 3D perception for robotic manipulation and object modeling, thus the capability of sensing at short distances is important. This is possible with both types of cameras.

ToF cameras can acquire images at 0.2m. At this distance, and considering the field of view, even relatively small objects, like a plant leaf, fill a large part of the image (Figs. 2a - 2c). Kinect minimum working distance is specified at 0.7m, but depth images can be obtained up to 0.5m. At this distance and considering the wider field of view, the same leaf fills only a small portion of the Kinect image. To permit the observation of details and comparison with ToF, Figs. 2d-2f show a cropped portion of the original Kinect images.

Getting closer to the object has two drawbacks for ToF cameras. On the one hand, focus problems appear (in practice this means a drop in the quality of the



(a) ToF depth (b) ToF intensity (c) ToF 3D point cloud

(d) Kinect depth (e) Kinect color (f) Kinect 3D point cloud

Fig. 2 Typical images supplied by a ToF camera and a Kinect camera at their shortest working distances. Original Kinect images are cropped to facilitate the comparison and observation of details. (a) Depth is codified as color. The details of the vein structure are observed, while in (d) they are not retained. (c) and (f) are the reconstructed 3D point clouds for each camera using factory settings. Observe in (c) the false *flying points* between the leaf edge and background, and in (d), the holes between the leaf and the background due to occlusions between the infrared (IR) light projector and the camera.

computed depth). 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 for close objects then the depth of field is small. ToF cameras do not have auto-focus capabilities, so the focus (and consequently the desired depth of field) has to be determined in advance. On the other hand, currently 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 particular range of depths. As before, for close distances the range of possible depths for a given integration time is small. Some ToF cameras have the capability of auto-adjusting the integration time. However, depth calibration of ToF cameras depends on integration time, and a common practice is to calibrate for only a few integration times, which are chosen considering the expected depth range.

Dense maps 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, occlusions). As will become apparent in the next sections, RGBD cameras are more sensitive to this problem by construction, as some points are visible by the camera and are occluded from the projector, and consequently their depths cannot be estimated. In practice this produces some discontinuities in the depth image, mainly at edges, represented as black zones (Fig. 2d).

Depth computation RGBD cameras do not directly compute the depth of image points. Instead, they compute first the disparity between the projected pattern points and the viewed ones. A careful calibration of the sensor is required to obtain precise depth values. The typical quantization problem of stereo systems appears also here, leading to an error in the depth measurements that increases quadratically with the distance from the sensor up to 4cm [13].

Camera model	PMD CamCube	Swissranger 4K	Kinect/Asus/Carmine 1.09
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	0.2 - 7m	0.8 - 5m 0.8 - 8m	0.7 - 3.5m 0.35-1.4m (Carmine 1.09)
Field of view	40x40	43.6x34.6 69x56	57x43
Focus	Adjustable	Fixed	Fixed
Integration time	Manual	Manual	Auto
Illumination	Auto	Auto	Auto (depth)
Outdoor	Suppression Background Illumination	Suppression Background Illumination	No
Images	Depth Intensity Amplitude Confidence	Depth Intensity Amplitude Confidence	Depth Color
Interface	USB	USB - Ethernet	USB

Table 1 Specifications of different ToF and RGBD cameras.

This effect, in the form of lack of details, can be observed in Figure 2d paying attention to the fact that the measured depths in the pixels of the whole leaf are almost the same. This can be produced by the interpolation process due to the special dotted pattern used as structured light, but also suggests, in conjunction with the lack of acquisition of small details (also shown in the next sections), that images delivered by Kinect are pre-processed with a smoothing filter, e.g. a Gaussian filter. In contrast, observe that the details of the vein structure are captured using ToF (Fig. 2a).

Classical stereo vision depth computation algorithms are a good alternative to obtain 3D maps. As it is known, they depend on the computation of disparities, from point features or image patches, that works better with textured surfaces. Even using global matching algorithms, accurate shape retrieval is hard when viewing untextured object surfaces like plant leaves [12]. A common technique is to use high resolution cameras (over 16 Mpixel) to ensure capturing enough texture. Such algorithms are costly in computation time, but GPU implementations can produce results in the order of one frame per second, which for some robotics applications may be adequate.

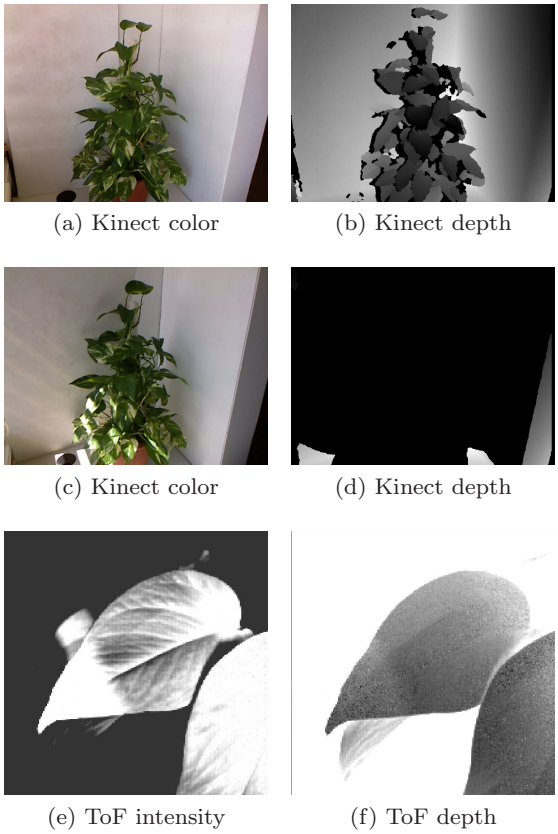
Colored point clouds One of the advantages of RGBD cameras is the ability to deliver colored depth points. The combination of ToF images and color images is also possible by computing the extrinsic calibration between both cameras [1] or alternatively using a beam splitter between the two cameras mounted at 90^o [9].

Illumination conditions All cameras can work in a wide variety of illumination conditions since all of them provide auto-illumination, except that Kinect cannot operate under strong lighting conditions like outdoors. Figure 3 shows an experiment where a plant is partially illuminated with direct sunlight, as it is common in greenhouses. Kinect was not designed to operate in these conditions, and we observe that in that scenario (Fig. 3c) it cannot provide depth information (Fig. 3d) while in dimmer light conditions it operates correctly (Figs. 3a and 3b). On the contrary (as shown in Figs. 3e and 3f), ToF cameras provide depth information but with noisier depth readings in those parts exposed to direct sunlight [12].

Calibration RGBD cameras use extrinsic parameters to correctly assign color to each depth point. In practice, calibration errors are easy to observe as coloring errors (Fig. 2f). Factory calibration parameters can be used or either standard calibration procedures can be applied [20].

Raw measurements captured by ToF cameras typically 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 remaining applications ToF cameras have to be calibrated over the specific application depth range [16]. A detailed description and classification of ToF errors can be found in [6].

A well-known problem of ToF images is the so called *flying points* (Fig. 2c). These are false points that ap-



(a) Kinect color (b) Kinect depth

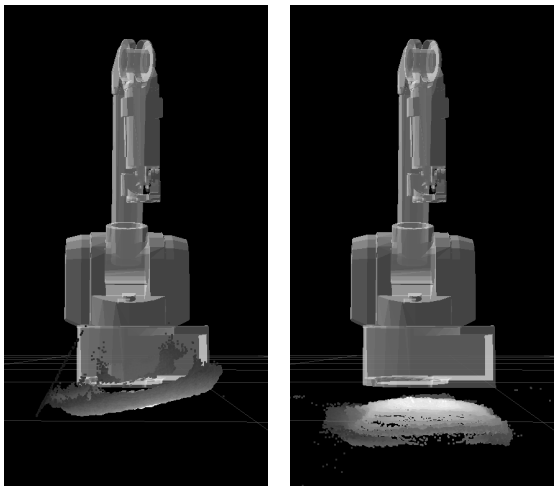
(c) Kinect color (d) Kinect depth

(e) ToF intensity (f) ToF depth

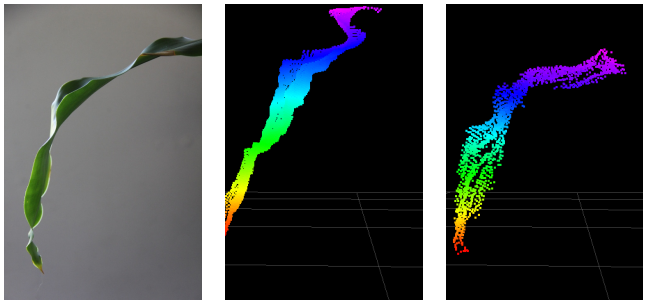
Fig. 3 Images in different sunlight conditions. (a),(b) Without direct sunlight, Kinect is capable of obtaining depth images. (c), (d) When parts of the plant receive direct sunlight (as it is common in greenhouses), Kinect cannot deliver depth information. (e), (f) ToF camera provides a depth image, even if sunlight partially illuminates a leaf. Observe, however, that overexposed leaf parts in the intensity image are noisier in the depth image.

pear between the edges of the objects and the background. These points have to be identified and filtered [17].

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 more susceptible to some error types but easier to calibrate [15]. Figure 4 exemplifies the benefits of a careful calibration. Observe that the 3D points in the calibrated image (Fig. 4b) correctly encode the shape of the real leaf (Fig. 4b). Special care should be taken to compensate errors due to saturation (amplitude-related) [5], light scattering [4] and multiple light reflections [8]. Note that newer ToF cameras allow to easily detect saturated pixels.



(a) Flat surface viewed from a robot: (left) uncalibrated, (right) calibrated.



(b) Complex leaf: (left) color image, (center) uncalibrated, (right) calibrated image correctly encodes the shape of the leaf.

Fig. 4 Comparison between uncalibrated and calibrated depth measurements.

3 Experimental assessment in three applications

We present images acquired in the three types of environment discussed in the Introduction, namely kitchen, botanic and textile, and provide hints to help select a camera depending on the particular demands of each task.

3.1 Object modeling in kitchen scenes

Cooking tableware is an example of small rigid objects, whose modeling plays a fundamental role prior to their manipulation by a robot.

Figures 5b and 5c show two examples taken with a CamCube camera and a Kinect, respectively. Observe that some details, like the edges of the three dishes and the scourer, can be identified in the ToF image and thus a correct manipulation action could be potentially triggered, but these details are not visible with Kinect.

Since their appearance in the market, ToF cameras have not been used extensively for object modeling because precise depth data is hard to obtain. Nevertheless, uncertainty reduction approaches can be used to mitigate error effects and make ToF cameras an adequate sensor. Algorithm 1 shows the main steps of an uncertainty reduction approach. The first part implements the idea of incrementally accumulating point clouds $\mathbf{S}_i$ acquired at different poses $\mathbf{T}_i$ using ICP. Up to this point, this approach suffers from the typical cumulative error. The novelty is to use the sensor's covariance $\mathbf{\Sigma}_{sensor}$ to propagate the uncertainty of ICP as well as to integrate the views using a Pose SLAM approach. The key advantage emerges as soon as a part of the model is viewed again. In this case, a loop is closed and a refined set of poses $\bar{\mathbf{T}}_i$ and covariances $\bar{\mathbf{\Sigma}}_{pose,i}$ can be computed that adequately distributes the cumulative error, thus yielding a more accurate model.

Algorithm 1 Multi-view modeling under uncertainty

$[\mathbf{S}_0, \mathbf{T}_0] \leftarrow$ Capture point cloud.
for $i = 1$ to number of poses **do**
  $[\mathbf{S}_i, \mathbf{T}_i] \leftarrow$ Capture point cloud.
  $[\mathbf{m}_i, \mathbf{T}_i] \leftarrow$ ICP registration $(\mathbf{S}_i, \mathbf{S}_{i-1}, \mathbf{T}_i)$
  $\mathbf{\Sigma}_{pose,i} \leftarrow$ ICP error propagation $(\mathbf{m}_i, \mathbf{\Sigma}_{sensor})$
  $\left[\bar{\mathbf{T}}_i, \bar{\mathbf{\Sigma}}_{pose,i}\right] \leftarrow$ Pose SLAM $(\mathbf{T}_i, \mathbf{\Sigma}_{pose,i})$
end for

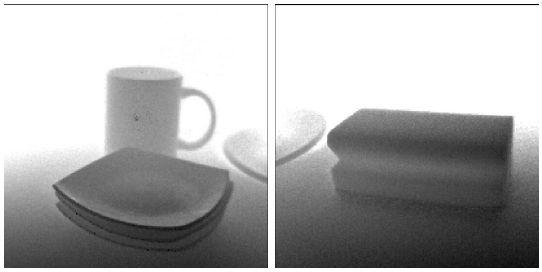
An example of such approach is the information-based SLAM method that we developed in the context of the PACO-PLUS project to improve 3-D point cloud registration [5]. Figure 5d shows an experiment to obtain a model of a water pitcher. The ToF camera is moved around the object and each new view is used to update the model using the uncertainty reduction approach. Videos of the experiments are available at `www.iri.upc.edu/groups/perception/activeSensing`.

3.2 Next-best-view planning for leaf measuring and cutting

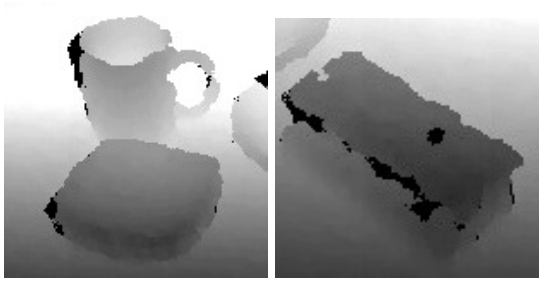
Algorithm 2 Plant leaf probing

$\mathcal{I} \leftarrow$ Move robot to initial position and get image
$L \leftarrow$ Leaves extraction $(\mathcal{I})$
repeat
  $p \leftarrow$ Select a target leaf (L)
  $\mathcal{I}_\nabla \leftarrow$ Move the robot to get better image (p)
  $L \leftarrow$ Leaves extraction $(\mathcal{I}_\nabla)$
  $G \leftarrow$ Extract grasping points (L)
  $l \leftarrow$ Detect target leaf(G)
until $\mathbf{g} \leftarrow$ suitable grasping point (l)
Sample leaf $(\mathbf{g})$

(a) Color image of the kitchen scene.



(b) ToF depth images taken at a close distance.

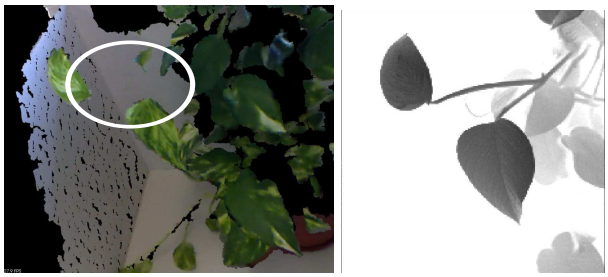


(c) Kinect depth detailed views (cropped).

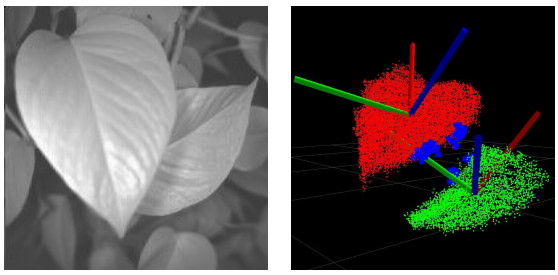
(d) Original water pitcher and its computed 3D model.

Fig. 5 Evaluation for kitchen objects. Depth is codified as grey value, where dark indicates short distances. (b) Using ToF, the edges of the different dishes can be observed and thus identified as stacked objects; also the foam scourer shape can be retrieved. (c) Kinect has difficulties sensing these details. (d) An uncertainty reduction approach is applied to obtain more accurate models using a ToF camera [5].

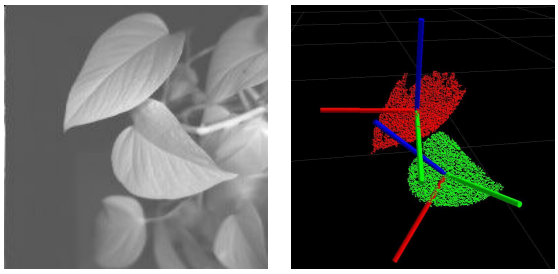
Precise depth data is required for robot applications involving contacts, like leaf sampling and chlorophyll measurement [1] and harvesting. Figure 6a shows that while Kinect is very good at obtaining the overall composition of the plants it misses some important structures, like branches (marked with a white ellipse). On the contrary, ToF correctly captures these details,



(a) Colored point cloud acquired with a Kinect (left) and ToF depth for the same plant portion (right).



(b) First view with ToF and 3D surface segmentation.



(c) Second view with ToF and 3D surface segmentation.

Fig. 6 (a) Kinect is very good at capturing the overall plant, but details such as branches, are not sensed. ToF correctly captures such details. (b) Plant leaves occlusions are common. (c) Sometimes it is possible to correctly model occluded leaves by moving the camera.

Algorithm 3 Leaves extraction ($\mathcal{I}$) [2]

1: Infrared-3D segmentation
2: Segment filtering
3: Construct segment graph representation
4: Graph-based clustering
5: Contour fitting
6: **return** $L \leftarrow$ List of leaves segments

even outdoors. Furthermore, infrared illumination reveals details like vein structure. In addition, the limited robot working-space imposes restrictions: given a desired point of view of the camera, a 3D vector pointing towards a point on the leaf, the robot can only displace the camera to a limited set of distances along this vector, usually the closest ones to the leaf. For example, using the robot-camera configuration shown in Fig. 1a, the 0.5m minimum distance of Kinect was a serious handicap as the robot could not reach most of the desired poses selected by the decision algorithm. Thus, sensors that operate at close distances are preferable.

Accordingly, within the project GARNICS we have worked on next view planning for plants using ToF cameras [2]. Our approach uses a combination of depth and infrared information to perform image segmentation and guidance of a robot equipped with tools for precise measurement and sample extraction. Algorithm 2 details the procedure we have proposed to acquire several images, from general to detailed, to obtain a leaf with the suitable characteristics and use the 3D information to perform the sampling task. Algorithm 3 sketches the vision algorithm that performs the segmentation of the different leaves present in an image. Registered infrared-3D data provided by the ToF camera is crucial in the segmentation and clustering steps. Figure 1 shows two examples of a robot using a custom cutting tool and an adapted chlorophyll meter with two different ToF cameras mounted in a hand-eye configuration.

Having the ability to move the camera is a key advantage in this context, as the modeling of plants is complex since plants have different shapes and details on some structures are important. Moreover, occlusions are very common. Figure 6 shows an example of such an advantage when a leaf is occluded by another leaf (Fig. 6b). Observe that in some situations the complete surface of the occluded leaf can be measured (Fig. 6c) by moving the camera to a new point of view. 3D information is very useful to segment leaves and determine edges, and is crucial for tasks requiring the alignment of sensors and tools in the surface of leaves, like the probing and measuring tasks tackled within the GARNICS project. Videos of the experiments are available at `www.iri.upc.edu/groups/perception/leafProbing`.

3.3 Textile object perception for manipulation

The manipulation of textiles is becoming a very active research topic due to its interest for service robotics as well as the availability of new dexterous manipulation tools. Figure 7a shows a close view of a folded shirt. Observe that ToF cameras offer good depth estimation of the shirt, and lots of details (even small wrinkles) can be identified. Similar results can be obtained using high resolution stereo cameras or ToF-color cameras combination [3]. On the contrary, Kinect detailed image reveals acquisition difficulties yielding some small holes in the surface. It should be noted that the position, size and number of holes (lack of data) vary while the sensor is moving.

In this application the ability to actively move the camera is not required, so a fixed Kinect camera looking at a table approximately 1 meter away was used. Additionally, Kinect offers color information without requiring additional calibration. Typical images acquired with the two sensors and an image obtained with this set-up are shown in Fig. 7, where the robot can also be seen from the camera point of view. Several algorithms have been applied that take advantage of 3D data in such context, like the detection of different kinds of wrinkles (Fig. 7b), and the identification of particular parts of clothing (Fig. 7c), like the collar of a polo shirt, or the construction of a wrinkledness map.

The experiments conducted [18] show that the depth data is informative enough and permits the analysis and extraction of different useful features to allow grasping. The training of a 3D descriptor to detect collars was performed using a blue polo, and the experiments included different pieces of clothing grouped in 3 sets: polo (only the blue polo, including slight and extreme deformations), mixed (blue polo appears mixed with other garments), and other (garments including polos but not the blue one). The comparison between using 2D methods alone or combined with 3D information shows that generalization (training only with a blue polo but testing with polos of other colors) is much better when 2D features are complemented with 3D information. Explanations of the different detectors and videos of the experiments are available at
`www.iri.upc.edu/groups/perception/wrinkledGrasping`.

4 Conclusions

Motivated by our involvement in robot manipulation projects requiring 3D object shape modeling (rather than recognition), we undertook a comparative performance assessment of 3D cameras, both RGBD and ToF ones. The main conclusion is that they exhibit complementary capabilities, i.e., the contexts for which one or the other camera seem more appropriate are different, and some applications would benefit from their combined use.

RGBD cameras don't require calibration and incorporate off-the-shelf procedures that make their usage easy and quick. Thus they are a good choice to readily get depth images of a scene. Their main shortcoming is that they are difficult to be used as active devices mounted on robots to work at short distance ranges. Moreover, the details that they can supply on the shape of objects are limited (see Figs. 5, 7 and specially Fig. 6), and they cannot operate outdoors.

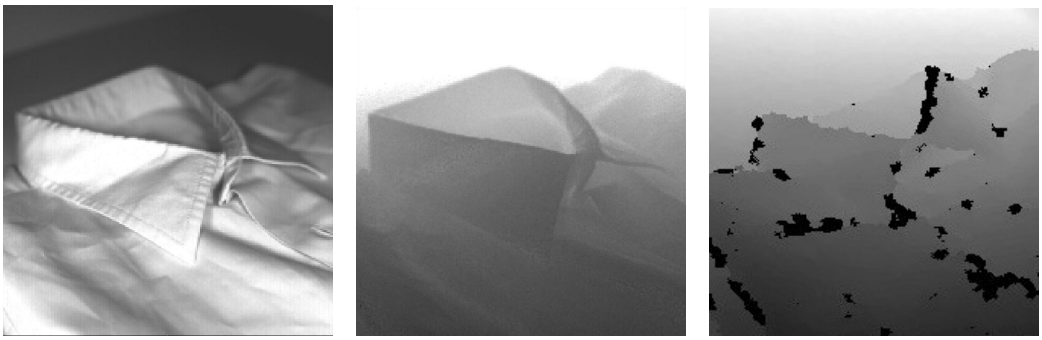
Although ToF cameras have apparently lower resolution, they can provide depth images at short distances of up to 20cm. This capability makes them very valuable in contexts where fine details on the objects are crucial. The price to pay is the need to manually set the focus, which determines the depth of field, as well as to tune the integration time, since each value yields good-quality images only for a narrow depth range. Depth calibration can also be performed to increase accuracy. Moreover, combining ToF and high-resolution color cameras requires additional calibration.

Another situation in which RGBD cameras have difficulties and ToF cameras do not is in providing depth images of scenes partially illuminated by extraneous light that superimposes on the projected light (see Fig. 3). A further issue to take into account are occlusions due to the separation between the light projection axis and the optical axis in the RGBD camera, and between the ToF sensor axis and the high-resolution camera axis when used in combination. This problem is avoided when using a ToF camera alone, since depth and intensity images are provided registered, and it can be surmounted in the other cases by appropriately placing a beam splitter.

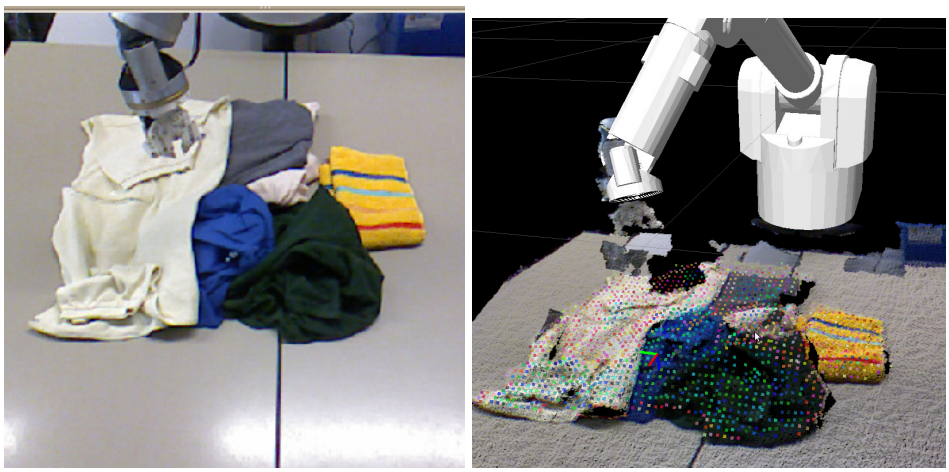
The conclusions of our assessment have guided the development of three applications, as briefly reported in the paper. In the PACO-PLUS project, we used a ToF camera under an uncertainty reduction SLAM approach to model rigid objects with curved shapes in a kitchen setting. Later, within the GARNICS project, we used a ToF camera under a next-best-view approach to find suitable leaves from which to take probes. Since this requires getting very close to the plant and finding suitable probing points with high precision, a ToF camera was more appropriate, although it required considerable parameter tuning. Finally, within the PAU+ project we have used a RGBD camera to develop perception algorithms for the manipulation of textile objects. RGBD cameras were preferred because they deliver colored depth points off-the-shelf and, compared to ToF, the camera does not need to be carefully calibrated to operate at different depth distances.

References

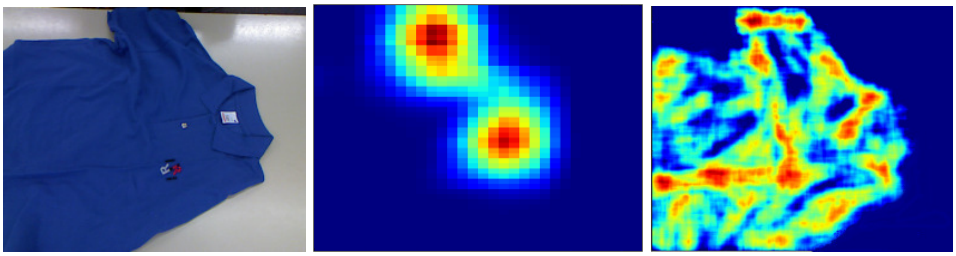
1. Alenyà G, Dellen B, Torras C (2011) 3D modelling of leaves from color and tof data for robotized plant measuring. In: Proc. IEEE Int. Conf. Robotics Autom., Shanghai, pp 3408–3414.
2. Alenyà G, Dellen B, Foix S., Torras C (2013) Robotized plant probing: Leaf segmentation utilizing time-of-flight data. IEEE Robotics and Automation Magazine, 20(3): 50–59.



(a) Folded shirt acquired with ToF ((left) intensity, (center) depth) and with Kinect ((right) cropped view to highlight details).



(b) (left) Wrinkled clothes on a table, and (right) their image taken with a Kinect with superimposed dots colored according to the 5 kinds of wrinkles detected, along with the 3D model of the robot.



(c) (left) Image of a polo, (center) two candidate regions to contain the collar, (right) corresponding wrinkledness map of the whole polo.

Fig. 7 Images of a folded shirt, a pile of clothings and a polo, together with the outputs of several methods that take advantage of 3D data for the perception of textiles: (b) wrinkle type detection, (b) collar detector and general wrinkledness.

3. Alenyà G, Foix S., Torras C (2014) ToF cameras for active vision in robotics. Sensors & Actuators: A. Physical , 218: 10–22.
4. Chiabrando F, Chiabrando R, Piatti, D, Rinaudo F (2009) Sensors for 3D imaging: metric evaluation and calibration of a CCD/CMOS time-of-flight camera. Sensors, 9(12), 10080–10096.
5. Foix S, Alenyà G, Andrade-Cetto J, Torras C (2010) Object modeling using a ToF camera under an uncertainty reduction approach. In: Proc. IEEE Int. Conf. Robotics Autom., Anchorage, pp 1306–1312.
6. Foix S, Alenyà G, Torras C (2011) Lock-in time-of-flight (ToF) cameras: A survey. IEEE Sensors J 11(9):1917–1926.
7. Fossati, A., Gall, J., Grabner, H., Ren, X., Konolige, K. (2013) Depth Cameras for Computer Vision. Advances in Computer Vision and Pattern Recognition Series. Springer.

8. Fuchs, S. and May, S. (2008) Calibration and registration for precise surface reconstruction with time of flight cameras, International Journal of Intelligent Systems Technologies and Applications 5(3/4):274–284.
9. Ghobadi S, Loepprich O, Ahmadov F, Bernshausen J. (2008) Real time hand based robot control using multimodal images Int. J. Comput. Science 35(4): 500–505.
10. Janoch, A., Karayev, S., Jia, Y., Barron, J. T., Fritz, M., Saenko, K., Darrell, T. (2013) A category-level 3d object dataset: Putting the Kinect to work. In Consumer Depth Cameras for Computer Vision, 141–165. Springer London.
11. Kahn, S., Bockholt, U., Kuijper, A., Fellner, D. W. (2013) Towards precise real-time 3D difference detection for industrial applications. Computers in Industry, 64(9), 1115–1128.
12. W. Kazmi, S. Foix, G. Alenyà, H.J. Andersen (2014) 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: 128–146.
13. Khoshelham K, Elberink SO (2012) Accuracy and resolution of Kinect depth data for indoor mapping applications. Sensors 12(2):1437–1454.
14. Lange R, Seitz P (2001) Solid-state time-of-flight range camera. IEEE J Quantum Electron 37(3):390–397.
15. Lefloch D, Nair R, Lenzen F, Schäfer H, Streeter L, Cree M, Koch R, Kolb A, Technical Foundation and Calibration Methods for Time-of-Flight Cameras, In M. Grzegorzek, C. Theobalt, R. Koch, A. Kolb (eds.), Time-of-Flight and Depth Imaging: Sensors, Algorithms, and Applications, LNCS 8200, Springer.
16. Lindner M, Schiller I, Kolb A, Koch R (2010) Time-of-flight sensor calibration for accurate range sensing. Comput Vis Image Underst 114(12):1318–1328.
17. May S, Fuchs S, Droeschel D, Holz D, Nuchter A (2009) Robust 3D-Mapping with Time-of-Flight cameras Proc. IEEE Int. Conf. in Intell. Robot Sys., Saint Louis, pp 1673–1679.
18. Ramisa A, Alenyà G, Moreno-Noguer F, Torras C (2012) Using depth and appearance features for informed robot grasping of highly wrinkled clothes. Proc. IEEE Int. Conf. Robotics Autom., Saint Paul, pp 1703–1708.
19. Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, Kipman A, Blake A (2011) Real-time human pose recognition in parts from a single depth image. Proc. 25th IEEE Conf. Comput. Vis. Pattern Recognit., Colorado Springs.
20. Zhang Z (2012) Microsoft Kinect sensor and its effect. IEEE Multimedia 19(2):4–10.