Taxonomy of 3D Sensors
A Survey of State-of-the-Art Consumer 3D-Reconstruction Sensors and their Field of Applications

Julius Schöning and Gunther Heidemann
Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

Keywords: 3D Sensors, Time of Flight, Structured Light, Taxonomy, 3D-reconstruction.

Abstract: Sensors used for 3D-reconstruction determine both the quality of the results and the nature of reconstruction algorithms. The spectrum of such sensors ranges from expensive to low cost, from highly specialized to out-of-the-shelf, and from stereo to mono sensors. The list of available sensors has been growing steadily and is becoming difficult to manage, even in the consumer sector. We provide a survey of existing consumer 3D sensors and a taxonomy for their assessment. This taxonomy provides information about recent developments, application domains and functional criteria. The focus of this survey is on low cost 3D sensors at an accessible price. Prototypes developed in academia are also very interesting, but the price of such sensors can not easily be estimated. We try to provide an unbiased basis for decision-making for specific 3D sensors.
In addition to the assessment of existing technologies, we provide a list of preferable features for 3D reconstruction sensors. We close with a discussion of common problems in available sensor systems and discuss common fields of application, as well as areas which could benefit from the application of such sensors.

1 INTRODUCTION

The first consumer RGB-D Camera named *Kinect* was launched in November 2010 by *Microsoft*. Before, RGB-D cameras were only available for specialized industrial applications. Triggered by this first low cost consumer RGB-D device, it was believed that a huge amount of consumer RGB-D cameras would be available on the market in the upcoming years and would provide a good alternative to, e.g., laser scanners. Now, six years later, the number of available RGB-D cameras is still limited. However, the available low cost RGB-D cameras are widely used in research for reconstruction (Handa et al., 2014; Cui et al., 2010), mapping (Henry et al., 2014; Huang et al., 2011), forensics (Dupuis et al., 2014; Nguyen et al., 2014), robotics (El-laithy et al., 2012; Yip et al., 2014) and various other applications (Banerjee et al., 2014; Gallo et al., 2011).

In this literature survey, we summarize and compare existing low cost 3D sensors. “Low cost” means a price below 5.000€. We try to verify the statement by Henry et al. (2014) that RGB-D cameras provide depth information only up to a limited distance of typically less than five meters. We introduce applications and point out some drawbacks of existing cameras. Therefore, a special focus on quality and nature of 3D-reconstruction algorithms and processes depending on these sensors is given. Finally, we discuss common problems in available sensors using a structured light camera in different test setups.

Contrary to initial expectations, academic prototypes like (Zollhöfer et al., 2014) cannot be taken into consideration in this taxonomy, because their total costs cannot be calculated without taking into account the work that has to be invested in order to set up such a prototype. Although the hardware costs may seem favorably small, total costs may well rise above 5.000€ if manpower is taken into account. Therefore, this survey discusses commercial, out-of-the-box systems only.

2 COMPARISON OF CONSUMER 3D SENSORS

All cameras discussed here can be assigned to one of two distinct groups according to their depth measurement principle: structured light (SL) or time of flight (ToF). Cameras working with structured light emit a light pattern onto the scene and calculate the depth

Schöning, J. and Heidemann, G.
Taxonomy of 3D Sensors - A Survey of State-of-the-Art Consumer 3D-Reconstruction Sensors and their Field of Applications.
DOI: 10.5220/0005784801920197
In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 3: VISAPP*, pages 194-199
ISBN: 978-989-758-175-5

Table 1: Comparison of consumer 3D-Cameras; *of measured distance, - not specified.

Camera	**Principle**	**measuring range [m]**	**Error**	**Res. RGB Res. Depth**	**FPS**	**FoV H×V**	**PL / SDK**	**Price [€]**
Structure Sensor	SL	$0,4-3,5$	1%*	640×480 640×480	30/60	$58^\circ\times45^\circ$	C/C++	305
Kinect 1st Gen.	SL	$0,4-3,5$	$<4cm$	640×480 640×480	15/30	$57^\circ\times43^\circ$	C#/C++/VB/ JAVA etc.	160
Xtion PRO Live	SL	$0,8-3,5$	-	1280×1024 640×480	30/60	$58^\circ\times45^\circ$	C#/C++/ JAVA	140
RealSense	SL	$0,2-1,2$	1%*	1920×1080 640×480	60	-	C#/C++/ JAVA/ JavaScript	80
Senz3D	ToF	$0,2-1,0$	-	1080×720 320×240	30	$74^\circ\times$ $41,6^\circ$	C++/C	115
Argos 3D - P100	ToF	$0,1-3,0$	<3 %*	n/a 160×120	160	$90^\circ\times$ $67,5^\circ$	Matlab/ Labview	1.200
Kinect 2nd Gen.	ToF	$0,5-4,5$	-	1920×1080 512×424	$15/30$	$70^\circ\times60^\circ$	C#/C++/C	160
Swiss Ranger 4500	ToF	$0,8-9,0$	$<4cm$	n/a 176×144	$10/30$	$69^\circ\times55^\circ$	C++/C/ Matlab/ Halcon	3.930
CamBoard pico	ToF	$0,2-1,0$	$<6mm$	n/a 160×120	45	$82^\circ\times66^\circ$	C++/C/ Matlab	585

information based on the deformation of the pattern. In most cases, the emitted light pattern has a wavelength in the infrared spectrum and is thus invisible for the human eye. ToF cameras use a laser to emit a pulse of light and calculate the distance based on the time span until the pulse is seen by the detector and the speed of light. Next to this main attribute, we use the field of view (FoV) characteristic, as shown in Figure 1, for defining the comparison of all sensors in Table 1. As an indicator for the fields of application, the SDKs and the supported programming languages (PL) for accessing the sensor software APIs are mentioned. The most important non-technical attribute of our comparison matrix is the price which allows us to calculate a price-performance ratio.

2.1 Structured Light Sensors

According to its fact sheet (Occipital, Inc, 2015), the *Structure Sensor* is designed to work at distances ranging from 40 centimeters to three and a half meters. The producer claims that the error in the z-direction (the direction of depth) is less than one percent of the actual distance. The camera is based on the structured light principle and provides a depth image with a resolution of 640×480 pixels at framerates between 30 and 60 frames per second (fps). However, its main drawback is its platform limitation since it is designed to work exclusively with an *Apple iPad*. This might potentially discourage *Windows* and *Linux* users, as well as users who require a more computationally powerful setup.

The first generation *Kinect* sensor (Microsoft, 2015b) is most prominent and well embraced by, e.g., the robotics community. It is built on the structured light principle as well. It features a depth range from 40 centimeters to $3,5$ meters, where the distance error under moderate constraints is below four centimeters (Khoshelham and Elberink, 2012). It offers the highest RGB-D resolution with 1280×1024 pixels at 15fps and a resolution of 640×480 at 30fps. Together with its low price, *Kinect* is an interesting option for benchmarking 3D-targeted research. However, since its maximum reliable distance of $3,5$ meter might be too small to cover all requirements for the envisioned field of application, it seems questionable whether it could fulfill the practical needs for applications which require reliability at a larger scale – however tempting it may seem at first.

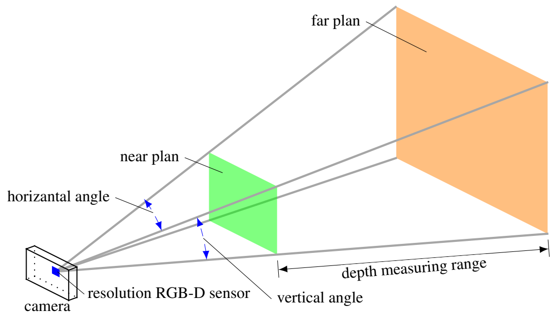


Figure 1: Field of view (FoV) characteristic incl. depth measuring range and resolution of the RGB-D sensor.

A sensor with specifications and capabilities comparable to the *Kinect* is the *Asus Xtion PRO Live* (ASUSTeK Computer Inc., 2015). Although the producer does not state the measurement principle, it is presumably also a structured light based method. At distances between 80 centimeters and $3,5$ meters it offers depth images at a resolution of 640×480 pixels. A fixed scanning frequency of 30fps is stated by the producer. However, with no declared error value, a slightly higher price and marginally lower resolution it offers no distinctive advantage over the *Kinect* sensor given the current application area.

In the first quarter of 2015, *Intels RealSense* camera was released. It targets the assessment of 3D point clouds at very small distances with a structured light depth sensor for distances between 20 centimeters and $1,2$ meters (Intel Corporation, 2015a,b). Its most salient advantages are its price, beating all other considered cameras by a margin of more than 80€, and the high framerate of 60fps.

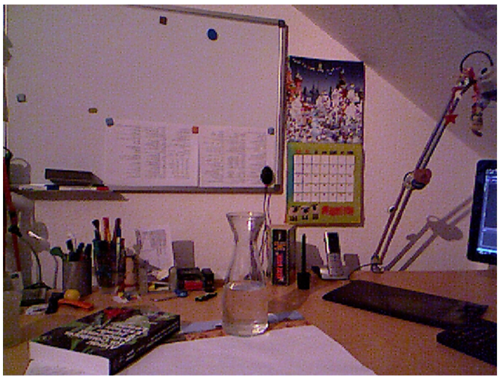
2.2 Time of Flight Sensors

A different type of camera uses the ToF physical measurement to assess the depth part of scenes.

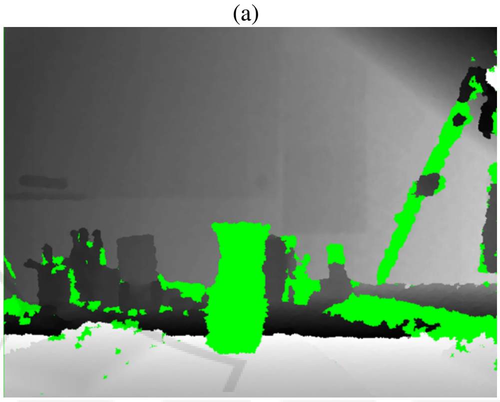
The *Creative Senz3D* (Creative Technology Ltd., 2015) is a ToF based depth camera with a targeted application area in human-computer-interaction (HCI). With an operating range between 20 centimeters and one meter it offers the coverage of a person located directly in front of a computer with the goal to recognize hand and arm gestures to be fed to an interface control task (cf. research area tangible user interfaces (TUI)). The resolution of the 3D depth images is limited to 320×240 pixels only, while the regular 2D webcam part of the sensor offers higher resolutions as well. Again, the manufacturer states no error values, which at 30fps and the lowest price of all considered cameras makes it an end user toy device not suitable for high quality application areas.

The *Argos 3D-P100* (Bluetechnix Group GmbH, 2015) creates depth measurements in a similar set of ranges as most of the cameras encountered before: Depth is measured between half a meter and three meters at an error rate below three percent. A resolution of 160×120 pixels is obtained at a framerate up to 160fps. Like other ToF cameras, the price of 1.200€ is above the consumer-eletronics level.

The second generation of *Microsoft Kinect* does not use structured light (Microsoft, 2015a) but relies on the ToF principle. Compared to the first generation, this results in a lower resolution of depth points (512 x 424 instead of 640 x 480) but also a slightly extended range of admissible depth values ($4,5$ instead of $3,5$ meters). Its most striking advantage, however, seems to be the increased horizontal and vertical viewing angle, giving the opportunity to obtain more overlapping regions in consecutive depth images.



(a)

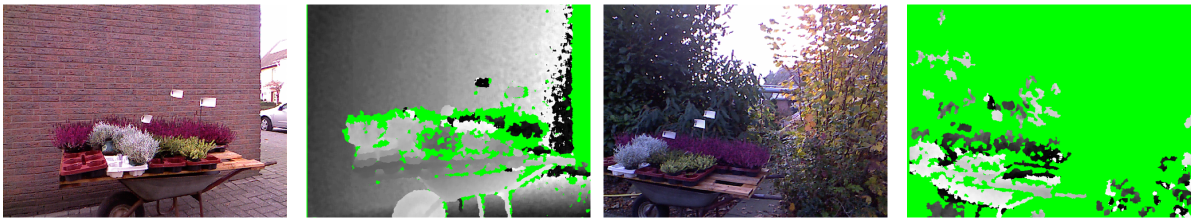


(b)

Figure 2: Indoor test scenario of an office desk with some objects; (a) shows the RGB channels and (b) shows the depth channel of the *Kinect* camera, where depth values above zero are represented as gray scale gradients.

The *Mesa Swiss Ranger (SR) 4500* ToF camera (Heptagon Micro Optics, 2015) offers a quite different depth sensing ability than the other sensors considered here. It can measure distances between 80 centimeters and nine meters with an error below four centimeters. The depth resolution of 176×144 pixels can be obtained between 10 and 30fps. Unfortunately, the price of nearly 4000 € marks the top end of the sensors and cameras considered in this survey.

Individual ToF camera modules can be assembled as, e.g., the *PMD CamBoard pico* (PMD Technologies GmbH, 2015). It offers depth measurements between 20 and 100 centimeters, but no error figures are provided. At resolutions of 160×120 pixels offered at 45fps, a three dimensional point cloud can be obtained. Nevertheless, due to the limited low range of measurement distances, application areas for this sensor seem limited.



(a) (b) (c) (d)

Figure 3: Two outdoor test scenarios, in the first scenario a wheelbarrow with plants without direct sun light; (a) shows the RGB channels and (b) shows the depth channel of the *Kinect* camera — in the second scenario a wheelbarrow with plants in front of trees with direct sun light; (c) shows the RGB channels and (d) shows the depth channel of the *Kinect* camera, where depth values above zero are represented as gray scale.

3 APPLICATION

An important question for typical research applications is whether the camera provides sufficient functionality and quality to perform sophisticated tasks like 3D reconstruction or mapping. Therefore, RGB and depth images in different scenarios will be analyzed. As a first test scenario, an indoor scene is evaluated using a structured light camera—the *Microsoft Kinect* of the first generation. Figure 2(a) shows the test scenario, a desk with ordinary objects like books, pencils, and input devices. On the right hand, in Figure 2(b) the eleven bit depth image is shown, where all areas marked green represent a depth value of zero, and depth values greater than zero are represented as grayscale gradient.

A 3D reconstruction application like, e.g., crime scene investigation might not be able to reconstruct this test scene because of missing depth information for some regions in the image. Notable hot spots, where the depth information does not correspond to the real depth, frequently occur on very smooth surfaces. Two of these spots are the transparent carafe in the center and the draw pad on the right-hand side. For 3D reconstruction approaches the depth information of transparent objects would be quite important because RGB based algorithms such as *structure from motion* cannot handle transparent objects well (Ihrke et al., 2010). The missing depth information for the following processes have to be handled by filling or filtering algorithms. One example is the voxel cloud connectivity segmentation (Papon et al., 2013), which uses the depth information next to RGB information for the segmentation process. In case of missing depth information, the voxel cloud connectivity segmentation performs hole filling using the SLIC algorithm. Another approach to handle missing depth information is the reconstruction of objects using inference from their depth-shadows (Albrecht and Marsland, 2013). Similar to reflections on smooth surfaces, in theory a further problem of cameras working with the structured light principle is the total absorption of the emitted (IR) light. However, in praxis we could not observe such an effect in a scenario with materials like fleece or velvet.

In the preface to the book "Consumer Depth Cameras for Computer Vision", Jamie Shotton argues that the depth camera technology is still not mature and has a long way to go to reach the frame rates and resolutions possible with traditional sensors, and does not yet work satisfactory outdoors (Fossati et al., 2013). To verify this statement about the outdoor capabilities of structured light cameras, we tested with two outdoor scenarios for 3D reconstruction. In the first outdoor scenario, Figure 3(a) and (b), the *Kinect* is used at cloudy weather conditions without direct sunlight. The depth information of Figure 3(b) is consistent and does not show abnormalities. With sunlight, as seen in Figures 3(c) and (d), the resulting depth information is greatly affected by the sunlight and of very limited use.

Compared to the structured light cameras, time-of-flight cameras show acceptable results in outdoor scenarios with or without sunlight. This has already been evaluated in agricultural robotic scenarios for orcharding and viticulture (Wunder et al., 2014). However, time-of-flight cameras still have a depth distance of nine meters only. This short depth distance limits the maximal driving speed of, e.g., autonomous vehicles because the inert drive train cannot stop instantly.

In 3D reconstruction, the limitation of the maximum depth also causes some drawbacks. With respect to the maximum depth distances, as shown in Figure 4, these low cost cameras can only scan small objects like persons or cupboards in a stationary setup. Using the camera as a hand scanner as in the *Kinect* fusion project (Newcombe et al., 2011; Microsoft Research, 2015) or similar approaches (Lee et al., 2014), the mentioned low cost 3D cameras yield good results. But if we are going to reconstruct large

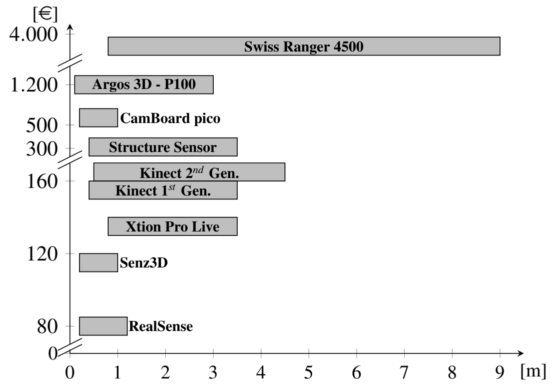


Figure 4: Comparison of consumer 3D-Cameras; price [€] over depth measuring range [m].

monuments (the size of Colognes famous Cathedral), the hand scanning device is clearly infeasible.

4 DISCUSSION AND OUTLOOK

Next to the price, an important reason for using a camera in a wide application field is the number of supported programming languages. For example, the Microsoft *Kinect* SDK of the first generation supports over four different programming languages. As a consequence, SDKs like OpenCV, ROS, PCL and OpenNI already offer implemented APIs for the *Kinect*. The powerful interfaces have led to success in diverse fields of application, which is remarkable for a camera originally designed for video gaming.

Figure 4 confirms the statement of Henry et al. (2014) that RGB-D cameras provide depth information only up to a limited distance of typically less than five meters. The *Mesa SR4500* with a depth of more than five meter is a depth only camera, thus it is not included in the statement of Henry et al. (2014). With a maximum reliable depth distance of around five meters, it appears questionable that requirements of outdoor applications can be fulfilled.

Structured Light Sensors are prone to very smooth or transparent objects, which lead to blind spots in the depth images. For 3D reconstruction in a controlled setup, such objects can be modified to achieve depth information. For example, DAVID Group (2015) offers a 3D coating spray, that can be applied to such objects for a reconstruction session and which is easy to remove. In a setup where very smooth or transparent objects such as windows can not be removed, the overlaying algorithms have to handle them, exclusively.

Summarizing, existing low cost cameras are a solid basis for indoor applications. But due to limited maximal depth and vulnerability by weather conditions (sunlight, fog etc.), outdoor usage remains very limited. Of course, such limitations can be cushioned by overlaying algorithms or special setups, but there are still challenging issues left. We noticed during this review that in particular the *Microsoft Kinect* camera is often regarded as a "professional" 3D camera, but its original purpose—video gaming—should still be kept in mind.

Depth information improves the 3D reconstruction process based on RGB data. But since depth information may not be available for all objects, it must be obtained from other sources. This is our motivation to develop an interactive approach (Schöning, 2015; Schöning and Heidemann, 2015), where the user can conveniently fill in missing depth information to achieve better 3D reconstruction and more reliable models.

ACKNOWLEDGEMENTS

This work was funded by the German Research Foundation (DFG) as part of the Scalable Visual Analytics Priority Program (SPP 1335).

REFERENCES

Albrecht, S. and Marsland, S. (2013). Seeing the unseen: Simple reconstruction of transparent objects from point cloud data. In *2nd Workshop on Robots in Clutter*.

ASUSTeK Computer Inc. (2015). Asus Xtion Specifications. http://www.asus.com/Multimedia/Xtion_PRO_LIVE/ specifications/.

Banerjee, T., Enayati, M., Keller, J. M., Skubic, M., Popescu, M., and Rantz, M. (2014). Monitoring patients in hospital beds using unobtrusive depth sensors. In *Engineering in Medicine and Biology Society (EMBC)*, pages 5904–5907.

Bluetechnix Group GmbH (2015). Argos 3D - P100 product website. http://www.bluetechnix.com/en/products/depthsensing/product/argos3d-p100/.

Creative Technology Ltd. (2015). Creative Senz3D website. http://us.creative.com/p/web-cameras/creative-senz3d.

Cui, Y., Schuon, S., Chan, D., Thrun, S., and Theobalt, C. (2010). 3D shape scanning with a time-of-flight camera. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1173–1180.

DAVID Group (2015). David - 3D coating spray. http://www.david-3d.com/products/accessories/coating-spray-500.

Dupuis, J., Paulus, S., Behmann, J., Plumer, L., and Kuhlmann, H. (2014). A multi-resolution approach for an automated fusion of different low-cost 3D sensors. *Sensors*, 14(4):7563–7579.

El-laithy, R., Huang, J., and Yeh, M. (2012). Study on the use of microsoft kinect for robotics applications. In *Position Location and Navigation Symposium (PLANS)*, pages 1280–1288.

Fossati, A., Gall, J., Grabner, H., Ren, X., and Konolige, K., editors (2013). *Consumer Depth Cameras for Computer Vision: Research Topics and Applications (Advances in Computer Vision and Pattern Recognition)*. Springer.

Gallo, L., Placitelli, A., and Ciampi, M. (2011). Controller-free exploration of medical image data: Experiencing the Kinect. In *Computer-Based Medical Systems (CBMS)*.

Handa, A., Whelan, T., McDonald, J., and Davison, A. J. (2014). A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM. *International Conference on Robotics and Automation*.

Henry, P., Krainin, M., Herbst, E., Ren, X., and Fox, D. (2014). RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments. In *Experimental Robotics*, pages 477–491. Springer.

Heptagon Micro Optics (2015). SR4500 data sheet. http://downloads.mesa-imaging.ch/dlm.php?fname=pdf/SR4500_DataSheet.pdf/.

Huang, A. S., Bachrach, A., Henry, P., Krainin, M., Fox, D., and Roy, N. (2011). Visual odometry and mapping for autonomous flight using an RGB-D camera. In *International Symposium of Robotics Research (ISRR)*.

Ihrke, I., Kutulakos, K. N., Lensch, H., Magnor, M., and Heidrich, W. (2010). Transparent and specular object reconstruction. In *Computer Graphics Forum*, volume 29, pages 2400–2426. Wiley Online Library.

Intel Corporation (2015a). Intel RealSense product brief. https://software.intel.com/sites/default/files/managed/0f/b0/IntelRealSense-WindowsSDKGold_PB_1114-FINAL.pdf.

Intel Corporation (2015b). RealSense 3D from lab to reality. http://iq-realsense.intel.com/from-lab-to-reality/.

Khoshelham, K. and Elberink, S. O. (2012). Accuracy and resolution of Kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454.

Lee, S.-O., Lim, H., Kim, H.-G., and Ahn, S. C. (2014). RGB-D fusion: Real-time robust tracking and dense mapping with RGB-D data fusion. In *Intelligent Robots and Systems (IROS)*, pages 2749–2754.

Microsoft (2015a). Kinect 2 for Windows technical datasheet. http://www.microsoft.com/en-us/kinectforwindows/meetkinect/features.aspx.

Microsoft (2015b). Kinect for Windows technical datasheet. https://readytogo.microsoft.com/en-us/_layouts/RTG/AssetViewer.aspx?AssetUrl=https%3A%2F%2Freadytogo.microsoft.com%2Fen-us%2FAsset%2FPages%2F08%20K4W%20Kinect%20for%20Windows_Technical%20Datasheet.aspx.

Microsoft Research (2015). 3D surface reconstruction. http://research.microsoft.com/en-us/projects/surfacerecon/.

Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., Kohi, P., Shotton, J., Hodges, S., and Fitzgibbon, A. (2011). KinectFusion: Real-time dense surface mapping and tracking. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 127–136.

Nguyen, T. V., Feng, J., and Yan, S. (2014). Seeing human weight from a single RGB-D image. *Journal of Computer Science and Technology*, 29(5):777–784.

Occipital, Inc (2015). Structure Sensor & SDK fact sheet. http://io.structure.assets.s3.amazonaws.com/Structure%20Sensor%20Press%20Kit.zip.

Papon, J., Abramov, A., Schoeler, M., and Wörgötter, F. (2013). Voxel cloud connectivity segmentation - supervoxels for point clouds. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2027–2034.

PMD Technologies GmbH (2015). Reference design brief CamBoard pico. http://www.pmdtec.com/html/pdf/PMD_RD_Brief_CB_pico_71.19k_V0103.pdf.

Schöning, J. (2015). Interactive 3D reconstruction: New opportunities for getting CAD-ready models. In *Imperial College Computing Student Workshop (ICCSW)*, volume 49 of *OpenAccess Series in Informatics (OASIcs)*, pages 54–61. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Schöning, J. and Heidemann, G. (2015). Interactive 3D modeling - a survey-based perspective on interactive 3D reconstruction. In *International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, volume 2, pages 289–294. SCITEPRESS.

Wunder, E., Linz, A., Ruckelshausen, A., and Trabhardt, A. (2014). Evaluation of 3D-sensorsystems for service robotics in orcharding and viticulture. In *VDI-Conference "Agricultural Engineering" VDI-Berichte Nr. 2226*, pages 83–88. VDI-Verlag GmbH Düsseldorf.

Yip, H. M., Ho, K. K., Chu, M., and Lai, K. (2014). Development of an omnidirectional mobile robot using a RGB-D sensor for indoor navigation. In *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, pages 162–167.

Zollhöfer, M., Theobalt, C., Stamminger, M., Nießner, M., Izadi, S., Rehmann, C., Zach, C., Fisher, M., Wu, C., Fitzgibbon, A., and et al. (2014). Real-time non-rigid reconstruction using an RGB-D camera. *ACM Transactions on Graphics*, 33(4):1–12.