By Georg Halmetschlager-Funek, Markus Suchi, Martin Kampel, and Markus Vincze

An Empirical Evaluation of Ten Depth Cameras

Bias, Precision, Lateral Noise, Different Lighting Conditions and Materials, and Multiple Sensor Setups in Indoor Environments

Over the last few years, novel color and depth sensors have pushed the boundaries of robot perception significantly. Today, several new depth-sensing products are replacing the earlier, well-examined red-green-blue-depth (RGBD) sensors, which have reached the end of their product life cycle and are no longer available. The properties of the new sensors have not yet been investigated, and it is unclear how they will compare to earlier RGBD sensors.

In this article, we evaluate ten different RGBD sensors that represent the three main sensor technologies: structured light, active stereo, and time of flight (ToF). Our work considers the influence of different target materials, different lighting conditions, and interference from other sensors in a multisensor setup. First, we collect 510 data points using ten different sensors in a robot setup to perform four experiments per sensor. We then evaluate the sensors by comparing five different metrics: bias, precision, lateral noise, behavior under different lighting conditions and materials, and the applicability of multiple sensor setups. Based on our results, we conclude with recommendations regarding which sensor to use for a given application.

RGBD Sensors

Since the 2010 introduction of Microsoft's affordable Kinect depth sensor [1], RGBD sensors have become an essential component in many methods and applications that use machine vision, especially in the field of robotics. RGBD sensors are used in robotics applications such as three-dimensional (3-D) simultaneous localization and mapping (SLAM) and navigation, reconstruction, object recognition

Digital Object Identifier 10.1109/MRA.2018.2852795
Date of publication: 13 August 2018

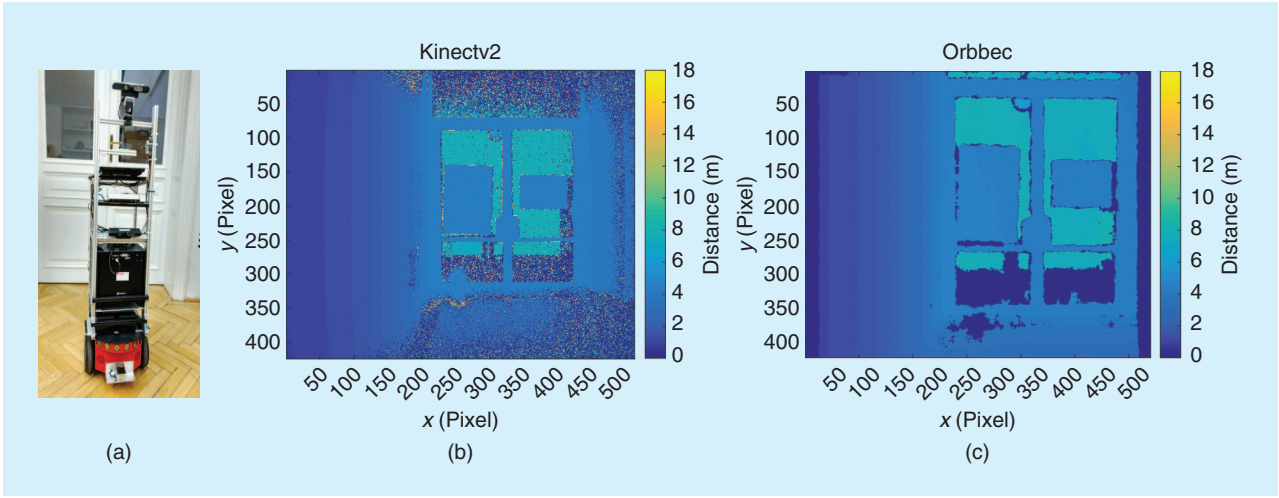

Figure 1. (a) A robot used for sensor integration. Depth data gathered with (b) a Kinectv2 sensor [5] and (c) an Orbbec 3-D sensor [6].

and tracking, human recognition and following, and hand gesture analysis. The Kinect and similar sensors provide not only two-dimensional (2-D) color data, but also depth measurements for each pixel. The transition from 2-D to 2.5-D opened up new opportunities for researchers to develop more sophisticated algorithms. Since then, a clear majority of leading research articles in the field of RGBD data processing exhibited a close relation to the Microsoft Kinect or its direct replacements, the Asus Xtion [2] or Primesense [3] RGBD sensor series.

The Kinect and similar sensors provide not only two-dimensional color data, but also depth measurements for each pixel.

Now, seven years after their release, those sensors have reached the end of the product life cycle. Fortunately, their popularity has led to several successor systems from different manufacturers that use similar or different technologies to provide RGBD data. New sensors using ToF or active stereo promise to offer similar or even improved depth-sensing results and different characteristics for robotics applications (Figure 1). However, sensor manufacturers provide no, or very limited, information regarding the sensors' noise behaviors [2], [4]–[12]. To our knowledge, a comprehensive overview that quantitatively evaluates these new sensor systems, including their noise characteristics in the robotics context, has not yet been published.

Many of the algorithms operating on RGBD data incorporate the specific noise characteristics of these sensors. For instance, SLAM algorithms include a model of the decreasing precision and bias of the measurements to determine the reliability of the measured data, which directly results in better performance [1]. The same applies, but is not limited to, object recognition, segmentation, 3-D reconstruction, and camera tracking [13]–[16]. Hence, knowing the correct noise models leads to better and more reliable results in various fields of robot vision.

This article contributes a comprehensive evaluation and comparison with respect to 1) different depth sensors and 2) different metrics. To be more precise, we analyze ten different sensors in terms of bias and precision as defined in [1], [17], and [18] under various conditions. We focus on indoor scenarios because sensors that rely on projected infrared patterns to obtain depth data are not designed to deal with incident sunlight. Our experiments are tailored to extend the results of [1] and provide a comprehensive and general overview without focusing on specific applications. Therefore, we designed several experiments to incorporate different distances, materials, and lighting conditions. We also investigate interference induced by other sensors, which is of special interest in robotic and multirobotic systems.

The ten sensors cover both near- and far-range devices, as well as three different sensor technologies: ToF, structured light, and active stereo. ToF sensors transmit light pulses to measure the time required for the pulses to travel from the light source to an object and back to the sensor. This duration determines the distance to the object. Structured light sensors project a (dot) pattern onto the observed surfaces and extract the depth information using the principal of triangulation. Active stereo cameras combine the idea of an active projector and a passive stereo camera pair. In contrast to classic stereo cameras, active stereo cameras additionally project their own texture. Thus, they are able to gather information even on low-textured surfaces.

Our comparison includes seven wide-range sensors—the Asus Xtion Pro Live [2], Orbbec Pro [5], Structure IO [4], Kinectv2 [6], RealSense D435 [7], RealSense ZR300 [8], and RealSense R200 [9]—as well as three near-range sensors: the RealSense SR300 [11], RealSense F200 [10], and Ensenso N35 [12]). Table 1 provides an overview of the key specifications of the sensors.

Table 1. The analyzed sensors and their specifications given by the manufacturers.

Sensor	Xtion Pro Live [2]	Structure IO [4]	Orbbec Pro [5]	Kinectv2 [6]	RS D435 [7]
Manufacturer	ASUS	Occipital, Inc.	Orbbec	Microsoft	Intel
Sensor type	RGBD	D	RGBD	RGBD	RGBD
Technology	Infrared pattern	Infrared pattern	Infrared pattern	ToF	Active stereo
Depth resolution	640 × 480	640 × 480	640 × 480	512 × 424	1,280 × 720
Range [m]	0.8–3.5	0.4–3.5+	0.6–8	0.5–4.5	0.2–10
Interface	USB 2	USB 2	USB 2	USB 3	USB 3
Sensor	**RS ZR300 [8]**	**RS R200 [9]**	**RS F200 [10]**	**RS SR300 [11]**	**Ensenso N35 [12]**
Manufacturer	Intel	Intel	Intel	Intel	Ensenso
Sensor type	RGBD	RGBD	RGBD	RGBD	D
Technology	Active stereo	Active stereo	Infrared pattern	Infrared pattern	Active stereo
Depth resolution	628 × 468	640 × 480	640 × 480	640 × 480	1,280 × 1,024
Range (m)	0.55–2.8	0.51–4	0.2–1.2	0.2–2	0.47–1.1
Interface	USB 3	USB 3	USB 3	USB 3	Ethernet

USB: universal serial bus.

A comprehensive statistical analysis using 40 experiments evaluating more than 50,000 images guarantees an impartial comparison of the different sensors. This work is focused on robotic-related machine vision systems; therefore, we integrate all sensors on a robot [Figure 1(a)] and evaluate the performance of the sensors using their standard configurations. This analysis is of particular interest for several robot perception tasks, i.e., those in which modeling the sensor noise increases the performance of the algorithms. This includes tasks for navigation, manipulation of objects that use RGBD reconstruction, and object detection and recognition.

Related Work

Nguyen et al. [1] investigated different noise characteristics of a Kinect sensor; to our knowledge, theirs is the most important work regarding RGBD sensor noise. The authors here proposed a 3-D noise distribution for Kinect depth measurements in terms of axial and lateral noise. Their work described in detail all of the experiments and metrics they used to quantize sensor noise in the context of reconstruction and tracking.

Han et al. [19] evaluated the potential of RGBD sensors for enhanced computer vision tasks. They reviewed the vision methods of data preprocessing, object tracking and recognition, human activity analysis, hand gesture analysis, and indoor 3-D mapping. They also considered the impact the Kinect RGBD sensor has had in terms of research and new technical challenges and demonstrated the importance of low-cost depth-sensing devices for the field of computer vision. Moreover, the article offers a short introduction to the technology used by RGBD sensors.

Andersen et al. provided a detailed analysis of the first Kinect sensor [20]. Their experiments used sequences of depth images, which allowed a statistical evaluation of bias, precision, resolution, influence of other sensors, and lateral noise. This approach was also used by Smisek et al. [21] and Pramerdorfer [22]. The main difference of [20] compared to our work is that the evaluation uses only an original Kinect sensor. Furthermore, the authors considered only one (or, at most, three) distances, depending on the experiment, and measured the influence of a single additional sensor. There was no evaluation of the sensor performance for different materials and lighting conditions.

An evaluation of sensor behavior with respect to multiple materials, including precision measurements on four materials, was performed by Berger et al. [23]. There were no other evaluation metrics presented in their work.

The idea of including different materials to evaluate Asus Xtion sensors for usability in fall-detection scenarios was also presented in the work of Pramerdorfer [22], which included a detailed description of the experiments conducted to evaluate resolution, lateral resolution, precision, sensor influence by adding one additional sensor, and bias. This work and [1] were the main inspirations for designing our own experiments. However, we

SLAM and reconstruction algorithms include a model of the decreasing precision and bias of the measurements to determine the reliability of the measured data and incorporate the noise characteristics.

extended the experiments with nine new sensors and six different materials and substituted the metric for lateral resolution in [22] with a similar measurement for lateral noise, as proposed in [1]. This resulted in a total of 510 data points for five different metrics.

Evaluation Metrics

This section describes the metrics used to evaluate the different sensors; keeping in mind robotics applications such as navigation, object detection, and human machine interaction, we propose five metrics. These metrics cover different aspects of noise and bias, reflection properties, the response to ambient illumination, and sensor interference. We first introduce the metrics and then outline how they capture the application requirements in an objective and measurable way.

We directly benchmark the sensors and not the underlying calibration methods; therefore, we use the raw sensor data together with the factory calibration.

Metrics 1 *and* 2, *bias and precision*: Bias [International Organization for Standardization (ISO) 5725-1: trueness] describes the deviation between the mean distance estimated by the sensor and the ground-truth distance. Precision quantizes the standard deviation of the depth measurements. Our definition of bias and precision follows the official definitions of trueness and precision according to ISO 5725-1 [24]. We use two discrete values to cover the full statistics of our measurement.

1) The bias is defined as

$$\text{bias} = \left| d_l - d_o - \mu_d \right|, \tag{1}$$

where d_l is the measurement of the laser device and d_o is the fixed distance offset between the mounted laser device and the tested sensor. Here, μ_d is the average depth defined as

$$\mu_d = \frac{1}{N \cdot n^2} \sum_{i=1}^{N} \sum_{u,v}^{n} I_i(u,v), \tag{2}$$

where N is the number of measurements, n is the size of the measured region, and (u, v) is the corresponding coordinate in the 2-D depth-image I_i.

2) The precision is defined as

$$\text{precision} = \sqrt{\left(\frac{1}{N \cdot n^2} \sum_{i=1}^{N} \sum_{u,v}^{n} \tilde{I}_i(u,v)^2 \right) - \tilde{\mu}_d^2}, \tag{3}$$

where $\tilde{I}_i$ is the corrected fronto-parallel depth image and $\tilde{\mu}_d$ is the average depth of $\tilde{I}_i$ using (2).

Metric 3, *lateral noise*: This metric quantizes the lateral noise around a vertical depth edge as a function of depth. We use the maximum distance of the image pixels around a depth edge to quantify the noise in pixels:

$$\text{latnoise}(d) = \arg\max_{p \in P} \left(\left| \Delta(p,l) \right| \right), \tag{4}$$

where p is an instance in the set of detected edge pixels P within a selected region (using Canny edges [25]), l is the least-mean-squares fitted line representing the edge, and $\Delta(.)$ is the pixel-line distance function.

The lateral noise may be transformed into a lateral resolution using a sensor's depth, lateral noise in pixels, and calibration parameters together with its projective geometry. In other words, this metric evaluates the precision of the sensor to quantize spatial expansions of objects and scenes in the image space (while the pixel value gives the depth measurement).

Metric 4, *lighting and materials*: This metric evaluates the precision depending on the reflectivity and absorption behavior of different materials in combination with the influence of ambient light. It indicates the performance of the sensor for different materials and under different lighting conditions.

Metric 5, *multiple sensors*: This metric quantizes the precision of a sensor in a multiple-sensor setup and the number of invalid values in relation to the full-sensor resolution (nan ratio). It is motivated by the fact that sensors using the same measurement technology tend to interfere with one another [11]. In other words, this metric measures the ability of the sensor to deal with multisensor setups, which occur in the field of robotics on a regular basis.

Why do these metrics capture the requirements of SLAM or reconstruction? SLAM and reconstruction algorithms include a model of the decreasing precision and bias of the measurements to determine the reliability of the measured data and incorporate the noise characteristics (incorporating a noise model results in better performance [1]). Incorporating the noise characteristics also improves the performance of the object recognition, segmentation, 3-D reconstruction, and camera tracking [13]–[16]. Although the main contribution of this article is an extensive evaluation of ten different sensors, we highlight the relevance of our results by showing the preliminary qualitative results of the reconstruction algorithm introduced in [26] incorporating our parametric error model.

Experimental Setup

Various publications are available regarding sensor calibration and depth offset compensation methods [27], [28]. We directly benchmark the sensors and not the underlying calibration methods; therefore, we use the raw sensor data together with the factory calibration. However, it should be noted that this article is primarily related to sensor calibration. Most calibration methods rely on information regarding the noise characteristics of raw sensor data to adapt the underlying noise and/or error model. This work provides the necessary information.

The data are gathered using a mobile platform equipped with the Robot Operating System (ROS) along with publicly available ROS wrappers for the sensors (Figure 2). The ground-truth measurements are obtained using a laser-range measurement device. Additionally, a luxmeter and a strong construction light are used to evaluate the sensors' capabilities under different lighting conditions.

Experiments 1 and 2: Bias, Precision, and Lateral Noise

Our gathered data for far-range sensors starts from the shortest distance at which the sensor is able to gather depth information and continues up to the farthest distance (~7 m) using a step size of 0.5 m. For near-range sensors, the measurements are conducted from approximately 0.3 m to 2 m with a step size of 0.1 m. The distances are validated using a laser measurement device. The depth offset between the sensor and the laser measurement device is determined manually and taken into account for the experiments. The sensor is positioned parallel to a planar surface. For each measurement, a region of 20 × 20 pixels on the target is recorded for 100 frames to make sure temporal noise is included in our evaluation. The ground-truth laser measurement is subtracted from the mean distance value obtained from the 100 frames to calculate the bias.

To determine precision, we fit a plane to the target area to compensate for the nonexact parallel alignment of the sensor to that area. This achieves a fronto-parallel sensor image. The obtained standard deviation gives a new data point for determining the precision of the sensor at the current distance.

For both bias and precision, we fit the parametric error model as follows:

$$f(d) = p_0 + p_1 \cdot d + p_2 \cdot d^2, \tag{5}$$

where d represents the depth and p_0, p_1, and p_2 are the coefficients of the quadratic error model. The determined error models for every sensor and the collected numerical data are publicly available on our webpage, https://www.acin.tuwien.ac.at/rgbd-sensor-tests/.

Similar to [1], we determine the lateral noise using the sharp vertical edge of the target. First, we manually select a region of the depth map containing the target's vertical edge. Second, we detect that edge using Canny edges [25] and fit a line model to the obtained pixels using least mean squares. This enables us to determine the distance of each edge pixel to the edge.

Experiment 3: Lighting and Materials

In this setup, six different materials under four different lighting conditions (4, 36, 277, and 535 lux) at distances 0.7 m (near range), 1 m, and 1.5 m (far range) are tested. The different lighting conditions are achieved by adding three light sources, one after the other, consisting of two ambient office lights and one strong spotlight.

The materials are chosen to cover a wide variety of reflective characteristics, including aluminum, black plastic, blue shiny plastic, foam, paper, and textile. The sensor is placed parallel to the objects. For each distance, object, and lighting condition, a region of 20 × 20 pixels is measured on the objects for 100 frames. The schematic of the experimental setup is depicted in Figure 3.

Experiment 4: Multiple Sensors

The simulation of interference by additional sensors is achieved by placing one additional sensor at a distance of 2 m and an angle of 60° and another at a distance of 1.1 m and an angle of 45° to the object (Figure 3). The interference measurements are conducted by adding one sensor after another. Each measurement, consisting of 100 frames, is taken from a planar surface parallel to the sensor.

Results

This section provides a comprehensive overview of the results achieved by our experiments, including interpretations and explanations.

Figure 2. The robotic system used for testing purposes.

Bias

While the Kinectv2 offers low bias over the whole range, we observe a significant increase of the bias for sensors using structured light starting from d>3 m. While all three structured light sensors and the two active stereo cameras (ZR300 and D435) offer a lower bias than the Kinectv2 for distances d<1 m, three sensors (ZR300, Orbbec, and Structure IO) offer an even lower bias for depth values d<2.5 m. We observe a quadratic increase of the bias for all sensors [full range: d = 0–8 m, Figure 4(a); zoom in: d = 0–3 m, Figure 4(b)]. The near-range sensors, F200 and SR300 [Figure 4(c)], show a slightly higher bias than their far-range counterparts, while the Ensenso N35 provides a low bias over the whole measurement range.

Precision

A quadratic decrease of precision is found in all far-range sensors [full range: d = 0–8 m, Figure 5(a); zoom in: d = 0–3, m, Figure 5(b)], but the structured light sensors differ in scale compared to the Kinectv2. Overall, the R200 and ZR300 sensors have the worst performance, while the Structure IO and Orbbec sensors perform very similarly. We observe that, at distances d<2 m, all structured light sensors generate less noisy measurements than the Kinectv2. Moreover, the D435 is able to gather more precise results than the Kinectv2 at distances d<1 m. We observe that the precision results for the D435 are more scattered than for the other sensors. The near-range sensors [Figure 5(c)] experience noise levels up to 0.0007 m. In the ranges specified by the manufacturers, we are able to obtain precision values under 0.004 m.

Lateral Noise

The analysis of lateral noise shows similar results for the three far-range structured light sensors and distances. For d<3 m, the noise level is independent of the distance, with three pixels for the structured light sensors and one for the Kinectv2 (Table 2). Two active stereo sensors (D435 and ZR300) offer a low lateral noise level similar to that of the Kinectv2. The R200 achieves a lower lateral noise of two pixels for distances closer than 2 m. In the near-range sensor, the Ensenso N35 achieves the highest lateral noise value.

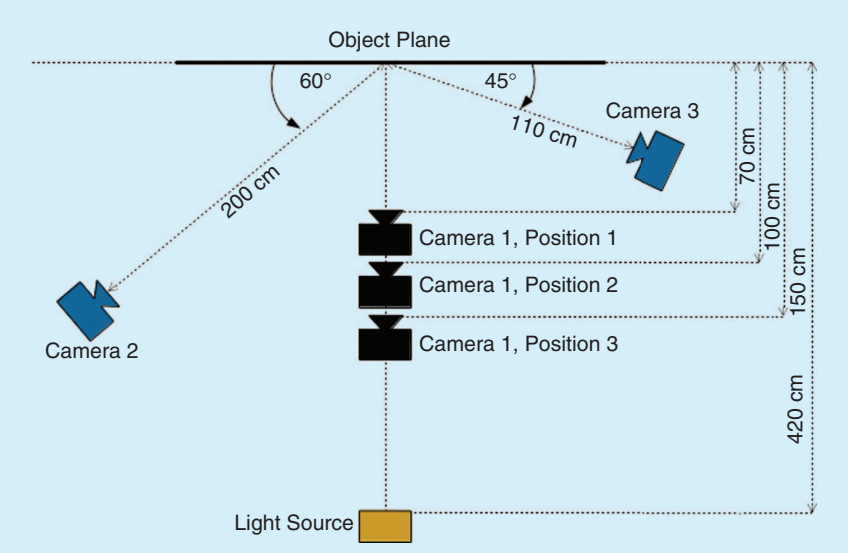

Figure 3. A setup for multisensor and material tests.

Materials

A total of 384 data points is gathered to determine how the sensors' precision is influenced by the reflection and absorption properties of six different materials in combination with four different lighting conditions from 4.2 to 535.75 lux (Figure 6). Figure 3 depicts the test setup.

The tests reveal that the Structure IO sensor best handles the varying object reflectances and lighting conditions. Although it has a lower precision compared to the other sensors for distances of d>1.5 m, it is able to

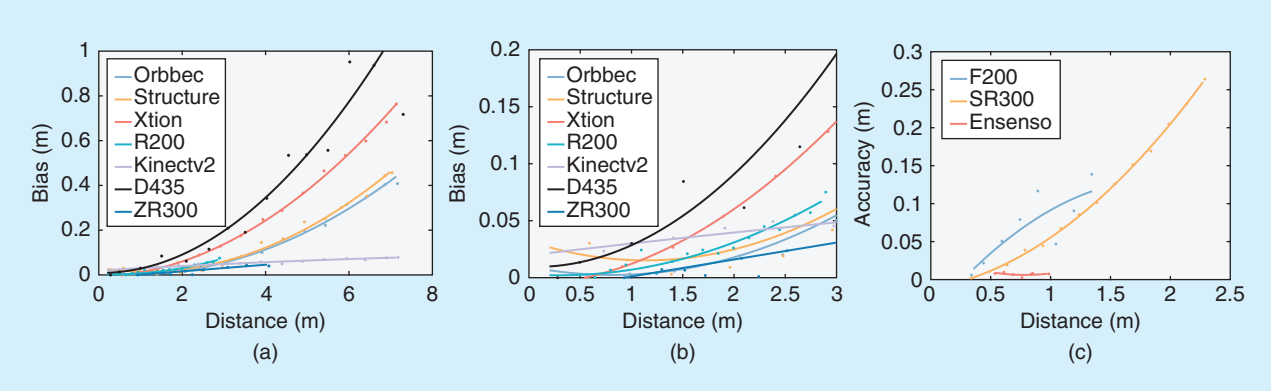

Figure 4. The results for bias (lower is better). (a) The bias of far-range sensors (d = 0–8 m), (b) the bias of far-range sensors (zoom in), and (c) the bias of near-range sensors.

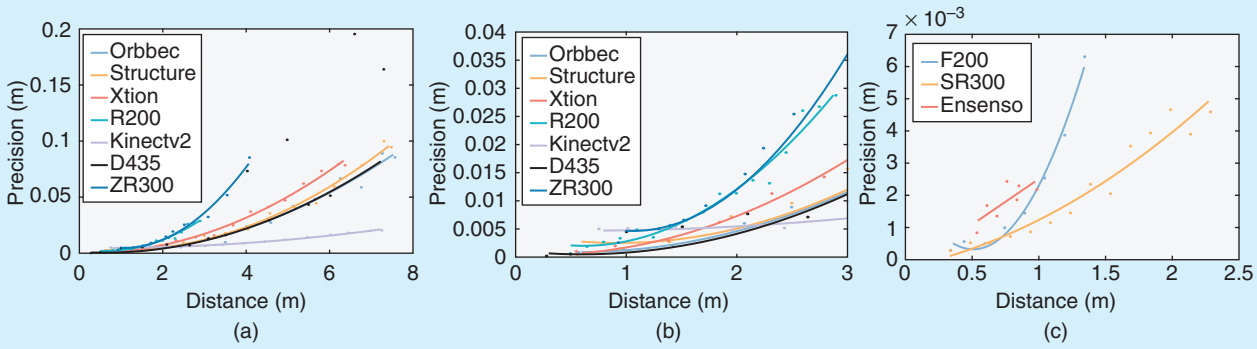

Figure 5. The precision results for near- and far-range devices (lower is better). (a) The precision of far-range sensors (d = 0–8 m), (b) the precision of far-range sensors (zoom in), and (c) the precision of near-range sensors.

gather information for high-reflective surfaces, such as aluminum, and under bright lighting conditions. While the Structure IO sensor gives a dense depth estimation, the Xtion is not able to determine a depth value. It is also notable that the Orbbec completely fails to gather depth information for four of the six surfaces under bright lighting conditions. The Kinectv2 fails to gather reliable depth data for aluminum at distances of d = 1 m and d = 1.5 m and under bright lighting conditions. The F200 and SR300 sensors have a significantly lower precision for bright lighting conditions. During the setup of the experiments, we expected the active stereo cameras (Ensenso and R200) to be able to handle different lighting conditions better than the structured light sensors due to the nature of their technology; this expectation was partially fulfilled.

Additional Sensors

Our results (Figure 7) reveal that the far-range structured light sensors can handle noise induced by one and two additional sensors. An exception occurs when the distance to the target is d = 1.5 m and two additional sensors are introduced to the scene. We did not observe a similar effect for the Kinectv2. The sensor gives stable results for precision independent of one or two additional sensors. The near-range sensors F200 and SR300 are significantly less precise with an additional sensor, and the Ensenso N35 is only slightly affected by a third observing sensor. At this point, we note that the high nan ratio for the close-range devices can be partially derived from our setup. Half of the scene is out of the sensor's range (Figure 8).

To summarize, the first experiment with one sensor provides a baseline for the measurements with two and three sensors observing the scene. The first differences are already visible if only one sensor is added. In particular, the SR300 and F200 sensors have a significant increase in the nan ratio if another Realsense device is added to the scene. For a closer analysis, we show the corresponding depth images. In Figure 8, it is clear that the depth extraction is heavily influenced by an additional sensor. The Ensenso and Kinectv2 sensors are nearly unaffected by the additional sensors.

Table 2. The lateral noise.

	Lateral Noise (Pixel)		
Sensor	**d = 0–0.7 m**	**d = 0.7–3 m**	**d = 3–5 m**
Asus Xtion	–	3	2
Structure IO	–	3	2
Orbbec 3-D	–	3	2
Kinectv2	–	1	1
D435	1	1	2
ZR300	–	0.5–1.2	–
R200	–	2–3	–
F200	1.5	–	–
SR300	2	–	–
Ensenso	3	–	–

Use Case

Schreiberhuber et al. [26] develop a scalable reconstruction method that uses a mesh to represent surfaces. In their work, they incorporate our error model for the precision of the RGBD sensor. As an outcome, they show a significant quality improvement of the reconstruction (Figure 9).

Discussion

The three far-range sensors using structured light show similar results for bias, precision, lateral noise, and noise induced by additional sensors. Their precision differs for different object properties and under varying lighting conditions. While the Structure IO sensor gathers valid depth data under all lighting conditions for all materials, it shows a slightly lower precision than the other sensors. The Orbbec sensor fails to gather data under bright lighting conditions for four of the six materials at a distance of 1 m. The difference in performance under bright lighting conditions may be related to the built-in infrared cameras, their dynamic range, and the performance of the auto exposure.

Figure 6. The precision ratios for different materials and lighting conditions (lower is better). A precision of zero indicates that the sensor is not able to gather any depth information.

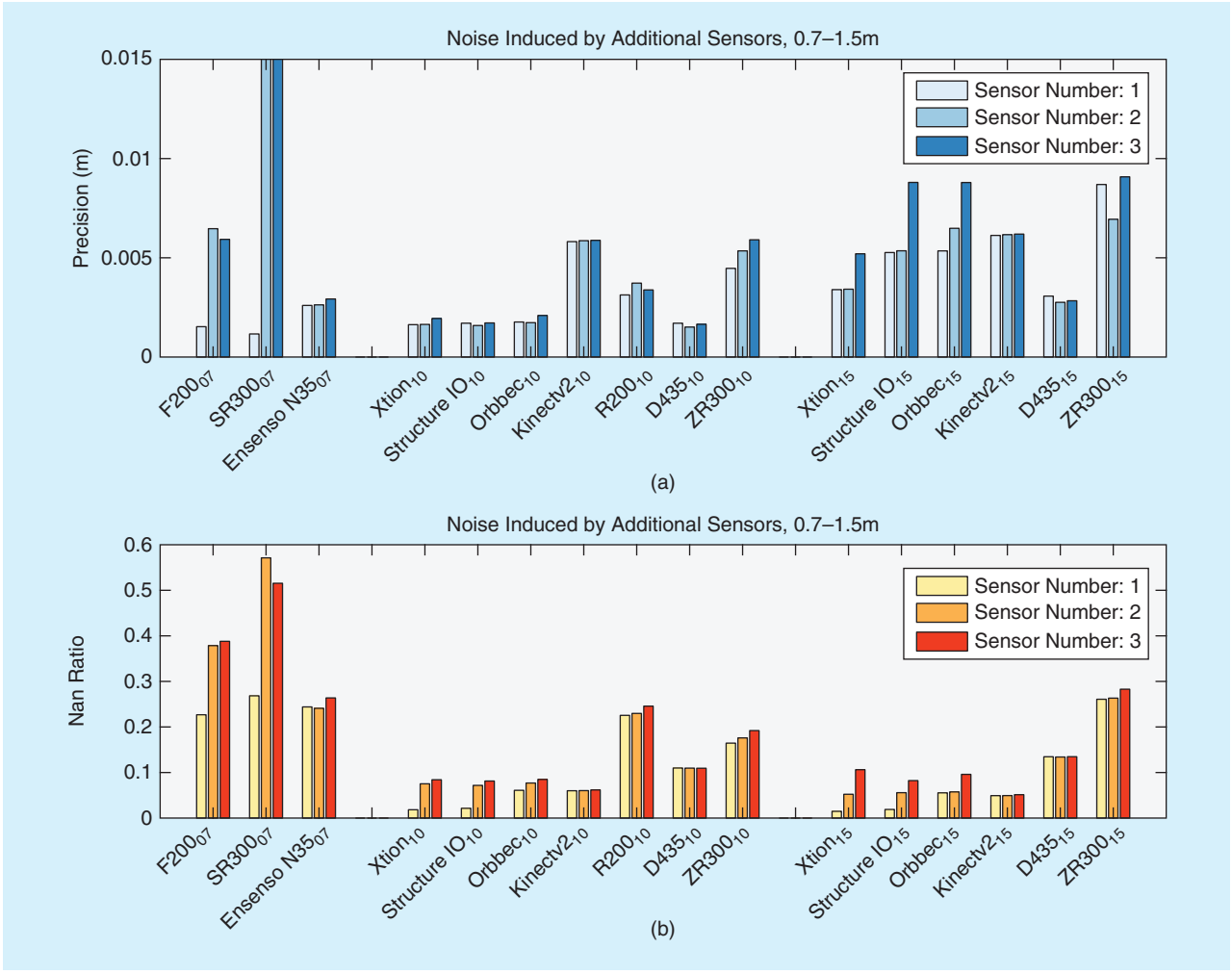

Figure 7. The (a) precision and (b) nan ratios (lower is better) in multisensor setups. The indices represent the distance to the target: 07 = 0.7 m, 10 = 1 m, and 15 = 1.5 m.

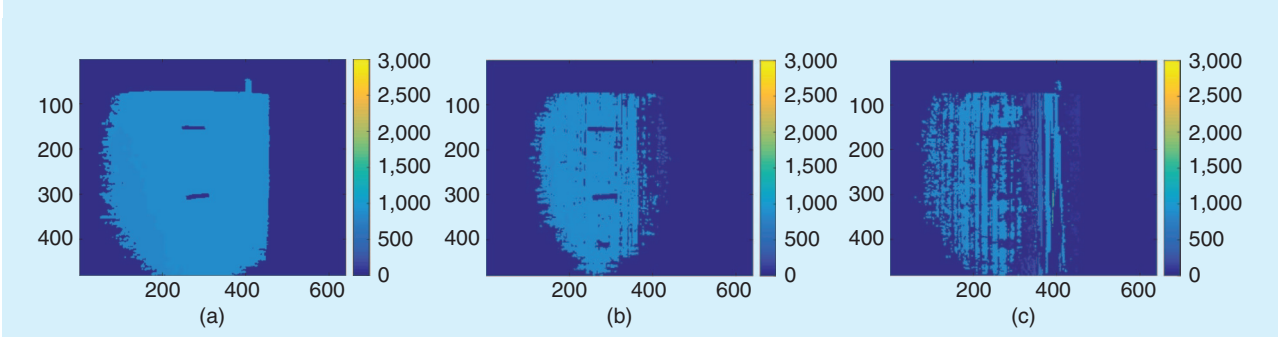

Figure 8. The influence of additional sensors (SR300). The number of sensors observing the scene: (a) 1, (b) 2, and (c) 3.

The RealSense R200 achieves a similar bias compared to the structured light sensors, while the ZR300 shows a smaller bias than the structured light sensors for d<2 m. However, the ZR300 appears to be less precise than the structured light sensors, independent of the target material. Moreover, they fail to gather depth data under bright lighting conditions.

The Microsoft Kinectv2 sensor behaves significantly differently compared to the other sensors. It outperforms all sensors regarding bias, lateral noise, and precision for d>2 m. For the range of 0.7 m<d<2 m, the Kinectv2 is less precise than the structured light sensors. Overall, the data provided by the Kinectv2 are less smooth and generate inferior surface representations for midrange depths.

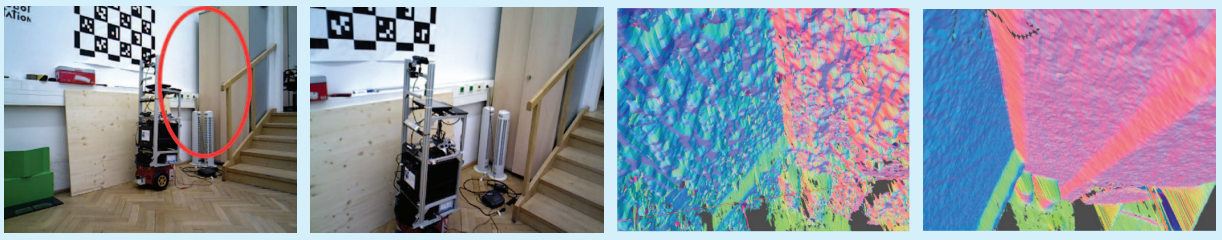

Figure 9. The reconstruction method from [26]. The two RGB images: (a) and (b) show the same location observed from different distances and (c) and (d) show the reconstruction results of the highlighted region (red circle), (c) without and (d) with our error model.

The D435 sensor gives scattered results for the precision and the bias. For some distances, it achieves a similar bias and precision as the Xtion sensor. However, for other distances we observe large outliers. Due to the nature of the used active stereo technology, the D435 is not influenced by additional sensors.

For the near-range devices, our experiments reveal that the F200 and SR300 sensors are not able to handle noise induced by additional sensors. Their precision and nan ratios are significantly influenced if a second sensor is added to the scene. In terms of precision, the F200 and SR300 are superior compared to the Ensenso N35 active stereo system. For all other metrics, the Ensenso N35 outperforms the two sensors.

Conclusions

This work evaluated ten different depth sensors using five metrics, aiming to achieve representative and comparable results to benchmark the sensors. Therefore, we semiautomatically collected 510 data points, each based on 100 depth frames. The results provided valuable information about state-of-the-art depth sensors for research in robotic perception and related applications.

Our investigation suggests the use of far-range structured light cameras for any application in which the quality of the surface representation is more relevant than the bias of the depth measures. This can include common robot tasks such as object modeling and recognition within the robot's manipulation distance, i.e., distances fewer than 2 m for approaching and handling an object. Moreover, the Asus Xtion and Structure IO sensors were able to gather data under all tested lighting conditions for all materials. Hence, they are especially useful for robots operating under uncontrolled conditions. Other comparisons revealed the following.

- The ZR300 sensor offered a low bias for <4 m but may fail to gather depth data under bright lighting conditions. Applications in which the bias was more relevant than the precision of the measurements fit the domain of this sensor.
- The D435 provided a remarkably wide range, from 0.2 to 7 m, and performed especially well for depth ranges <1 m. However, it failed during our experiments to gather depth measurements under bright lighting conditions and had scattered results for precision and bias.
- For large distances >4 m, the tested ToF sensor (Kinectv2) gathered the most reliable measurements, even under bright lighting conditions.
- The Ensenso active stereo camera offered the lowest bias within its narrow range of 0.5–1 m. It satisfied applications requiring low-biased measurements from a sensor that can be used out of the box.
- The RealSense ZR300, R200, and D435 sensors offered various parameters for adapting the sensor properties to the scene. During our experiments, we used the factory presets for high-accuracy measurements without adaptation to the current lighting and materials.

Because experiments 1 and 2 (bias and precision) are easy to reproduce, we suggest that users with new sensors gather data in a similar way to 1) benchmark their sensor against our results and 2) apply the previously introduced easy-to-use error model.

Future work should evaluate the different sensors under outdoor lighting conditions and add new sensors, as they are continuously entering the market. Furthermore, the experiments regarding precision, bias, and lateral noise could be extended for different viewing angles. The setup may be modified by replacing standard drivers with more advanced methods. For example, [29] implements a method to calculate a disparity map for structured light sensors, and [30] adds a filter to the Freenect2 driver to extend the sensor range, which may be of relevance to robotics applications.

Acknowledgments

This work was partially supported through the Horizon 2020 Program, grant number 519625 FLOBOT; funding from the Austrian Science Fund, grant number I1856-N30 ALOOF; and by the Vienna Science and Technology Fund, grant number ICT15-045 RALLI. We would like to thank Farhoud Malekghasemi and Simon Schreiberhuber for helping with their experiments. We are also grateful to CogVis GmbH (http://www.cogvis.at/) for providing a Structure IO sensor for the experiments.

References

[1] C. V. Nguyen, S. Izadi, and D. Lovell, "Modeling Kinect sensor noise for improved 3-D reconstruction and tracking," in *Proc. Second Int. Conf. 3-D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, 2012, pp. 524–530.

[2] ASUS. (2017). Specifications Xtion pro live. ASUS. [Online]. Available: https://www.asus.com/3D-Sensor/Xtion_PRO_LIVE/specifications/

[3] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, "Depth mapping using projected patterns," U.S. Patent 8150142, Apr. 3, 2012.

[4] Structure. (2018). Precise 3D vision for embedded applications. [Online]. Available: https://structure.io/embedded

[5] Orbbec. (2018). Orbbec Astra Pro. [Online]. Available: https://orbbec3d.com/product-astra-pro/

[6] Microsoft. (2017). Kinect for Windows. [Online]. Available: https://developer.microsoft.com/en-us/windows/kinect/hardware

[7] Intel Corporation. (2018). Intel RealSense camera D435 retrieved 2018-06-28. [Online]. Available: https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-D400-Series-Datasheet.pdf

[8] Intel Corporation. (2018). Intel RealSense Camera ZR300. Intel Corp. Santa Clara, CA. [Online]. Available: https://click.intel.com/media/ZR300-Product-Datasheet-Public-002.pdf

[9] Intel Corporation. (2018). Intel RealSense Camera R200. Intel Corp. Santa Clara, CA. [Online]. Available: https://software.intel.com/sites/default/files/managed/d7/a9/realsense-camera-r200-product-datasheet.pdf

[10] Intel Corporation. (2018). Intel RealSense Camera F200. Intel Corp. Santa Clara, CA. [Online]. Available: https://communities.intel.com/docs/DOC-24012

[11] Intel Corporation. (2018). Intel RealSense Camera SR300. Intel Corp. Santa Clara, CA. [Online]. Available: https://software.intel.com/sites/default/files/managed/0c/ec/realsense-sr300-product-datasheet-rev-1-0.pdf

[12] Ensenso. (2018). Specifications N35 Series. Ensenso Freiburg, Germany. [Online]. Available: https://www.ensenso.com/support/modellisting/?id=N35-804-16-IR

[13] T. Fäulhammer, R. Ambruş, C. Burbridge, M. Zillich, J. Folkesson, N. Hawes, P. Jensfelt, and M. Vincze, "Autonomous learning of object models on a mobile robot," *IEEE Robot. Autom. Lett.*, vol. 2, no. 1, pp. 26–33, 2017.

[14] K. Tateno, F. Tombari, and N. Navab, "Real-time and scalable incremental segmentation on dense SLAM," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, 2015, pp. 4465–4472.

[15] M. Nießner, M. Zollhöfer, S. Izadi, and M. Stamminger, "Real-time 3D reconstruction at scale using voxel hashing," *ACM Trans. Graph. (TOG)*, vol. 32, no. 6, pp. 169:1–169:11, 2013.

[16] Q.-Y. Zhou and V. Koltun, "Depth camera tracking with contour cues," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 632–638.

[17] Bureau International des Poids et Mesures. (2008). International vocabulary of metrology–Basic and general concepts and associated terms. *JCGM*. [Online]. *200*. Available: https://www.bipm.org/en/publications/guides/vim

[18] B. N. Taylor and C. E. Kuyatt. (1994). Guidelines for evaluating and expressing the uncertainty of NIST measurement results. U.S. Dept. Commerce, Technol. Admin., Nat. Inst. Standards and Technol. Gaithersburg, MD. [Online]. Available: https://emtoolbox.nist.gov/Publications/NISTTechnicalNote1297s.pdf

[19] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with Microsoft Kinect sensor: A review," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, 2013.

[20] M. R. Andersen, T. Jensen, P. Lisouski, A. K. Mortensen, M. K. Hansen, T. Gregersen, and P. Ahrendt, "Kinect depth sensor evaluation for computer vision applications," Dept. Eng., Aarhus Univ., Denmark, Tech. Rep. ECE-TR-6, 2012.

[21] J. Smisek, M. Jancosek, and T. Pajdla, "3D with Kinect," in *Consumer Depth Cameras for Computer Vision*. London: Springer-Verlag, 2013, pp. 3–25.

[22] C. Pramerdorfer, "Depth data analysis for fall detection," M.S. thesis, Dept. Comput. Sci., TU Wien, Austria, 2013.

[23] K. Berger, K. Ruhl, C. Brümmer, Y. Schröder, A. Scholz, and M. Magnor, "Markerless motion capture using multiple color-depth sensors," in *Proc. 16th Annu. Workshop Vision, Modeling and Visualization (VMV)*, 2011, pp. 317–324.

[24] *Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 1: General Principles and Definitions*, ISO 5725-1:1994.

[25] J. Canny, "A computational approach to edge detection," in *Readings in Computer Vision*. New York: Elsevier, 1987, pp. 184–203.

[26] S. Schreiberhuber, J. Prankl, and M. Vincze, "Scalablefusion 0.5: Feasibility analysis of a mesh based 3D reconstruction approach," in *Proc. OAGM Workshop*, 2018, pp. 47–52.

[27] A. Teichman, S. Miller, and S. Thrun, "Unsupervised intrinsic calibration of depth sensors via SLAM," in *Proc. Robotics: Science and Systems*, 2013.

[28] B. Zeisl and M. Pollefeys, "Structure-based auto-calibration of RGB-D sensors," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2016, pp. 5076–5083.

[29] S. Ryan Fanello, C. Rhemann, V. Tankovich, A. Kowdle, S. Orts Escolano, D. Kim, and S. Izadi, "Hyperdepth: Learning depth from structured light without matching," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 5441–5450.

[30] F. J. Lawin, P.-E. Forssén, and H. Ovrén, "Efficient multi-frequency phase unwrapping using kernel density estimation," in *European Conference on Computer Vision (ECCV)*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Berlin, Germany: Springer-Verlag, 2016, pp. 170–185.

Georg Halmetschlager-Funek (co-first author), Vision for Robotics Laboratory, Automation and Control Institute, Technische Universität Wien, Vienna, Austria. E-mail: gh@acin.tuwien.ac.at.

Markus Suchi (co-first author), Vision for Robotics Laboratory, Automation and Control Institute, Technische Universität Wien, Vienna, Austria. E-mail: suchi@acin.tuwien.ac.at.

Martin Kampel, Computer Vision Lab, Technische Universität Wien, Vienna, Austria. E-mail: martin.kampel@tuwien.ac.at.

Markus Vincze, Vision for Robotics Laboratory, Automation and Control Institute, Technische Universität Wien, Vienna, Austria. E-mail: vincze@acin.tuwien.ac.at.