An Empirical Evaluation of Ten Depth Cameras for Indoor Environments

Georg Halmetschlager-Funek[1,*] Markus Suchi[1,*] Martin Kampel[2] Markus Vincze[1]

***Abstract*— In the last few years novel color and depth sensors greatly pushed the frontier of robot perception. Now, several new depth sensing products are replacing the well examined first RGBD sensors that reached the end of the product life cycle and are no longer available. The properties of the new sensors have not yet been investigated in research and it is unclear how they compare to the first RGBD sensors. In this work we evaluate ten different RGBD and depth sensors covering the main three sensor technologies: Structured Light, Active Stereo, and Time of Flight. Our work evaluates the influence of different target materials, different lighting conditions, and interference from other sensors in a multi-sensor setup. We collect 510 data points using ten different sensors in a robot setup to perform in total 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, as well as the applicability for multiple sensor setups. Based on our results we conclude with recommendations for which sensor to use for which application.**

I. Introduction

Since the introduction of Microsoft's affordable depth-sensor, the Kinect [1], in the year 2010, RGBD sensors have become an essential component for many methods and applications using machine vision, especially in the field of robotics. RGBD sensors are used in robotics applications such as 3D SLAM and navigation, reconstruction, object recognition and tracking, human recognition and following, and hand gesture analysis. The Kinect and similar sensors provide not only 2D color data, but also depth measurements for each pixel. The transition from 2D to "2.5D" opened up new opportunities for researchers to develop more sophisticated algorithms. A clear majority of leading research articles in the field of RGBD data processing are developed in close relation to the Microsoft Kinect or its direct substitutors: the Asus Xtion [2] or Primesense [3] RGBD sensor series.

Now, seven years after their release, those sensors reached the end of the product life cycle. Fortunately, their popularity has led to several successor systems from different manufactures that use similar or different technologies to provide RGBD data. New sensors using Time of Flight (ToF) or Active Stereo promise to offer similar or even better depth-sensing results and different characteristics for robotics applications (cf. Figure 1). However, the sensor manufactures provide no, or only limited, information regarding the sensors' noise behaviors [2], [4]–[12]. To our knowledge a comprehensive overview that quantitatively evaluates the new sensor systems, including their noise characteristics in the robotics context, is still missing.

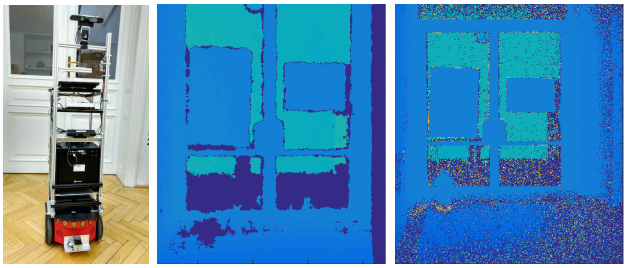


Fig. 1: Left: Robot used for sensor integration. Middle/Right: Depth data gathered with different Sensors (Middle: Orbbec 3D[5], Right: Kinectv2[6]).

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 accounts for, but is not limited to, object recognition, segmentation, 3D reconstruction, and camera tracking [13]–[16]. Hence, knowing the correct noise models lead to better and more reliable results in various fields of robot vision.

This work contributes a comprehensive evaluation and comparison with respect to i) different depth sensors and ii) different metrics. To be more precise, we analyze ten different sensors in terms of bias and precision as defined in [1], [17], [18] under various conditions. We focus on indoor scenarios since 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 to give a comprehensive and general overview without focusing on certain applications. Hence, several experiments are designed to incorporate different distances, materials, and lighting conditions. We also investigate interference induced by other sensors, which is of special interest in robotic and multi-robotic 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 in order to measure the time the pulses need to travel from

*G. Halmetschlager-Funek and M. Suchi are co-first authors.

[1]G. Halmetschlager-Funek, M. Suchi, M. Vincze are with the Vision for Robotics Laboratory, Automation and Control Institute, TU Wien, 1040 Vienna, Austria. {gh,msu,vm}@acin.tuwien.ac.at

[2]M. Kampel is with the Computer Vision Lab, TU Wien, Favoritenstr. 9-11/183-2, 1040 Wien, Austria. martin.kampel@tuwien.ac.at

This work was partially supported through the Horizon 2020 Programme, grant No. 519625 Flobot, and funding from the Austrian Science Foundation for grant No. I1856-N30 ALOOF.

the light source, to an object, and back to the sensor. Based on this duration, the distance to the object is determined. 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. Hence, they are able to gather information even on low textured surfaces.

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

A comprehensive statistical analysis using 40 experiments, evaluating over 50000 images guarantees an impartial comparison of the different sensors. This work is focused on robotic related machine vision systems, hence we integrate all sensors on a robot (cf. Figure 1) and evaluate the performance of the sensors using their standard configurations. This analysis is of high interest for several robot perception tasks, i.e., where modeling the sensor noise increases the performance of the algorithms. This includes tasks for navigation, manipulation of objects that use RGBD reconstruction, as well as object detection and recognition.

II. Related Work

Nguyen et al. [1] investigate different noise characteristics of a Kinect sensor and to our knowledge it is the most important and considered work regarding RGBD sensor noise. They propose a 3D noise distribution for Kinect depth measurements in terms of axial and lateral noise. Their work describes in detail all experiments and metrics they use to quantize sensor noise in the context of reconstruction and tracking.

Han et al. [19] evaluate the potential of RGBD sensors for enhanced computer vision tasks. They review the vision methods of data preprocessing, object tracking and recognition, human activity analysis, hand gesture analysis, and indoor 3D mapping. They consider the impact the Kinect RGBD sensor has had in research and new technical challenges opened up by this sensor. It shows the importance of low cost depth sensing devices for the field of Computer Vision. Moreover it gives a short introduction to the technology used by the RGBD sensors.

Andersen et al. provide a detailed analysis of the first Kinect sensor [20]. The conducted experiments use sequences of depth images, allowing a statistical evaluation of bias, precision, resolution, influence of other sensors, and lateral noise. This approach is also used by Smisek et al. in [21] and Pramerdorfer [22]. The main difference to our work is that the evaluation uses only an original Kinect sensor. Furthermore, they only consider one or at most three distances depending on the experiment, and measure the influence of a single additional sensor. There is no evaluation of the sensor performance for different materials and lighting conditions.

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

The idea of including different materials was also used in the work of Pramerdorfer [22], which evaluated Asus Xtion Sensors for usability in fall detection scenarios. The work includes 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 input for designing our own experiments. However, we extend the experiments with nine new sensors, six different materials, and substitute the metric for lateral resolution of [22] with a similar measurement for lateral noise as proposed in [1]. This results in overall 510 data points for five different metrics.

III. Evaluation Metrics

This section describes the metrics used to evaluate the different sensors. Following the review of related work and 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.

Metrics I and II - Bias & Precision – Bias (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 these two discrete values to cover the full statistics of our measurement:

$$bias = \mid d^l - d^o - \mu_d \mid, \tag{1}$$

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. 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, (u, v) is the corresponding coordinate in the 2D depth-image I_i. The precision is defined as follows:

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

TABLE I: 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 3D	Microsoft	Intel
Sensor Type	RGBD	D	RGBD	RGBD	RGBD
Technology	Infrared pattern	Infrared pattern	Infrared pattern	Time of Flight	Active Stereo
Resolution	640x480 (Depth)	640x480 (Depth)	640x480 (Depth)	512x424 (Depth)	1280x720 (Depth)
Range [m]	0.8 to 3.5	0.4 to 3.5+	0.6 to 8.0	0.5 to 4.5	0.2 to 10
Interface	USB 2.0	USB 2.0	USB 2.0	USB 3.0	USB 3.0

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
Resolution	628x468	640x480 (Depth)	640x480 (Depth)	640x480 (Depth)	1280 x 1024 (Depth)
Range [m]	0.55-2.8	0.51 to 4.0	0.2 to 1.2	0.2 to 2.0	0.47 to 1.1
Interface	USB 3.0	USB 3.0	USB 3.0	USB 3.0	Ethernet

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 III - 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:

$$latnoise = \arg\max_{p \in P}(\mid \Delta(p, l) \mid), \tag{4}$$

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-squares fitted line representing the edge, and $\Delta(.)$ is the pixel-line distance function.

The lateral noise may be transformed into a lateral resolution by using the depth, the lateral noise in pixels, and the calibration parameters of a sensor 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 IV - Lighting & Materials – This metric evaluates the precision depending on the reflectivity and absorption behavior of different materials in combination with the influence of ambient light. This metric indicates the performance of the sensor for different materials and under different lighting conditions.

Metric V - 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). This is motivated by the fact that sensors using the same measurement technology tend to interfere with each other [11]. In other words this metric measures the capability of the sensor to deal with multi-sensor setups occurring in the field of robotics on a regular basis.

Why do the metrics capture the requirements of e.g. 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 to incorporate the noise characteristics. It has been shown that incorporating a noise model results in better performance [1]. The same accounts for, but is not limited to, object recognition, segmentation, 3D reconstruction, and camera tracking [13]–[16]. Although the main contribution of this paper is an extensive evaluation of ten different sensors, we highlight the relevance of our results by showing preliminary qualitative results of the reconstruction algorithm introduced in [26] that incorporates our parametric error model.

IV. Experimental Setup

This section describes the setup of the experiments and the method used to produce results for the different metrics introduced in the previous section.

There exist various publications 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 work is highly 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 was gathered using a mobile platform equipped with ROS and the publicly available ROS wrappers for the used sensors. (cf. Figure 2). The ground truth measurements were obtained using a laser range measurement device. In addition, a luxmeter and a strong construction light was used to evaluate the sensors' capabilities under different lighting conditions.

A. Experiment I/II - Bias, Precision & Lateral Noise

Gathering data for far-range sensors is done starting from the shortest distance at which the sensor is able to gather depth information up to the furthest distance ($\sim 7.0m$) using a step size of $0.5m$. For near-range sensors, measurements are conducted from approximately $0.3m$ to $2.0m$ with a step size of $0.1m$. 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

Fig. 2: Robotic system.

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 that 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.

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

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

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

where d represents the depth and p_0, p_1, p_2 are the coefficients of the quadratic error model.

The determined error models for every sensor and the collected numerical data will be publicly available on our webpage[1].

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 vertical edge of the target. 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.

[1] https://www.acin.tuwien.ac.at/rgbd-sensor-tests/

B. Experiment III - Lighting & Materials

In this setup, six different materials under four different lighting conditions ($4lux$, $36lux$, $277lux$, $535lux$), at distances $0.7m$ (near-range), $1.0m$, and $1.5m$ (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. This includes 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.

C. Experiment IV - Multiple Sensors

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

V. Results

This section gives a comprehensive overview of the results achieved by our experiments including interpretation and explanations.

A. Precision

A quadratic decrease of precision is found in all far-range sensors (Full range: $d = 0 - 8m$, Figure 4a; Zoom-in: $d = 0m - 3m$, Figure 4b), but the Structured Light sensors differ in scale compared to Kinectv2. Overall the R200 and the ZR300 sensors have the worst performance, while the Structure IO and Orbbec sensors perform very similar. We observe that at distances $d < 2.0m$ all Structured Light sensors generate less noisy measurements than the Kinectv2. Moreover, the D435 was able to gather more precise results than the Kinectv2 at distances $d < 1m$. We observe that the precision results are more scattered than for the other

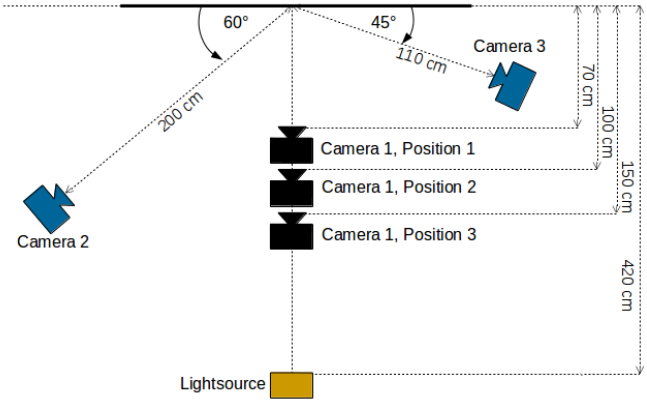


Fig. 3: Setup for multi-sensor and material tests.

TABLE II: Lateral Noise.

	Lateral Noise [pixel]		
Sensor	**d=0..0.7m**	**d=0.7..3m**	**d=3..5m**
Asus Xtion	-	3	2
Structure IO	-	3	2
Orbbec	-	3	2
Kinectv2	-	1	1
D435	1	1	2
ZR300	-	0.5-1.2	-
R200	-	2-3	-
F200	1.5	-	-
SR300	2	-	-
Ensenso	3	-	-

sensors. The near-range sensors (cf. Figure 4c) experience noise levels up to $0.007m$. In the ranges specified by the manufacturers we were able to obtain precision values under $0.004m$.

B. Bias

While the Kinectv2 offers a low bias over the whole range, we observe a significant increase of the bias for sensors using Structured Light starting from $d > 3m$. While all three Structured Light sensors and the two Active Stereo cameras ZR300 and D435 offer a lower bias than the Kinectv2 for distances $d < 1m$, three sensors (ZR300, Orbbec, Structure IO) offer an even lower bias for depth values $d < 2.5m$. We observe a quadratic increase of the bias for all sensors (Full range: $d = 0 - 8m$, Figure 5a; Zoom-in: $d = 0m - 3m$, Figure 5b). The near-range sensors F200 and SR300 (cf. Figure 5c) show a slightly higher bias than their far-range counterparts, while Ensenso N35 provides a low bias over the whole measurement range.

C. Lateral Noise

The analysis of lateral noise shows similar results for the three far-range Structured Light sensors and distances. For $d < 3m$ the noise level is independent of the distance, with 3 pixels for the Structured Light Sensors and 1 pixel for the Kinectv2 (cf. Table II). Two Active Stereo Sensors (D435 and ZR300) offer a similar low lateral noise level as the Kinectv2. The R200 achieves a lower lateral noise of 2 pixels for distances closer than $2m$. In the near-range the Ensenso N35 achieves the highest lateral noise value.

D. Materials

A total of 384 data points are gathered to determine how the precision of the sensors is influenced by the reflection and absorption properties of six different materials in combination with four different lighting conditions from $4.2lux$ to $535.75lux$ (cf. Figure 9). Figure 3 depicts the test setup.

The tests reveal that the Structure IO sensor handles the varying object reflectances and lighting conditions the best. Although it has a lower precision compared to the other sensors for distances $d > 1.5m$, it is able to 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 out of six surfaces under bright lighting conditions. The Kinectv2 fails to gather reliable depth data for aluminum at a distance of $d = 1m$ and $d = 1.5m$ under bright lighting conditions. The F200 and SR300 have a significant lower precision for bright lighting conditions. During the setup of the experiments we expected the Active Stereo cameras (Ensenso, 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.

E. Additional Sensors

Our results (cf. Figure 6) reveal that the far-range Structured Light sensors can handle noise induced by one and two additional sensors. One exception is when the distance to the target is $d = 1.5m$ 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 the precision independent of one or two additional sensors. The near-range sensors F200 and SR300 are significantly less precise with an additional sensor. The Ensenso N35 is only slightly affected by a third observing sensor. At this point we have to 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 (cf. Figure 7).

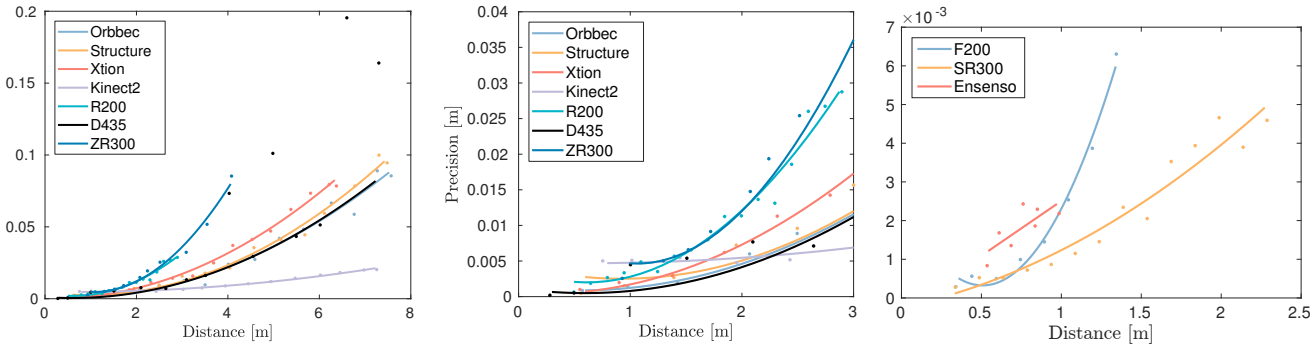
In summary, the first experiment with one sensor provides a baseline for the measurements with two and three sensors observing the scene. The first differences are already visible if only one sensor is added. In particular, the SR300 and F200 have a significant increase in the nan ratio if another RealSense device is added to the scene. For a closer analysis we show the corresponding depth images. In Figure 7 it is clear that the depth extraction is heavily influenced by an additional sensor. The Ensenso and Kinectv2 are nearly unaffected by the additional sensors.

F. 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 (cf. Figure 8).

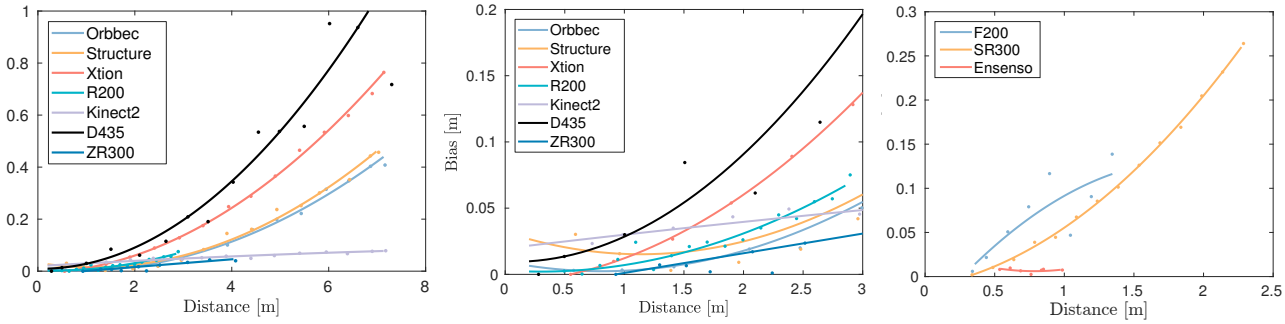
VI. Discussion

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 out of six materials at a distance of $1m$. The difference in performance under bright lighting conditions may be related to the built-in IR cameras, their dynamic range, and the performance of the auto exposure.



(a) Precision of far-range sensors. d=0-8m. (b) Precision of far-range sensors, zoom-in. (c) Precision of near-range sensors.

Fig. 4: Precision results for near and far-range devices. Lower is better.



(a) Bias of far-range sensors, d=0-8m. (b) Bias of far-range sensors, zoom-in. (c) Bias of near-range sensors.

Fig. 5: Results for bias. Lower is better.

The RealSense R200 achieves a similar bias than the Structured Light sensors, the ZR300 shows a smaller bias than the Structured Light sensors for $d < 2m$. 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 behaves significantly different compared to the other sensors. The Kinectv2 outperforms all sensors regarding bias, lateral noise, and precision for $d > 2m$. For the range of $0.7m < d < 2m$ the Kinectv2 is less precise than the Structured Light sensors. To conclude, the Kinectv2 is less smooth and generates worse surface representations for mid range depths.

The D435 gives scattered results for the precision and the bias. For some distances, the sensor achieves a similar bias and precision as the Xtion. 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 ratio 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.

VII. Conclusion

This work evaluates ten different depth sensors using five metrics, aiming to achieve representative and comparable results to benchmark the sensors. Therefore, we semi-automatically collected 510 data points, each based on 100 depth frames. The results provide 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 where 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 manipulation distance of a robot, i.e., distances below $2m$ for approaching and handling an object. Moreover the Asus Xtion and Structure IO Sensor were able to gather data under all tested lighting conditions for all materials. Hence, they are especially useful for robots operating under uncontrolled conditions.

The ZR300 offers a low bias for $< 4m$ but may fail to gather depth data under bright lighting conditions. Applications where the bias is more relevant than the precision of the measurements fit the domain of this sensor.

The D435 provides a remarkable wide range from $0.2m$ to $7m$ and performs especially well for depth ranges $< 1m$. However, it failed during our experiments to gather depth measurements under bright lighting conditions and had scat-

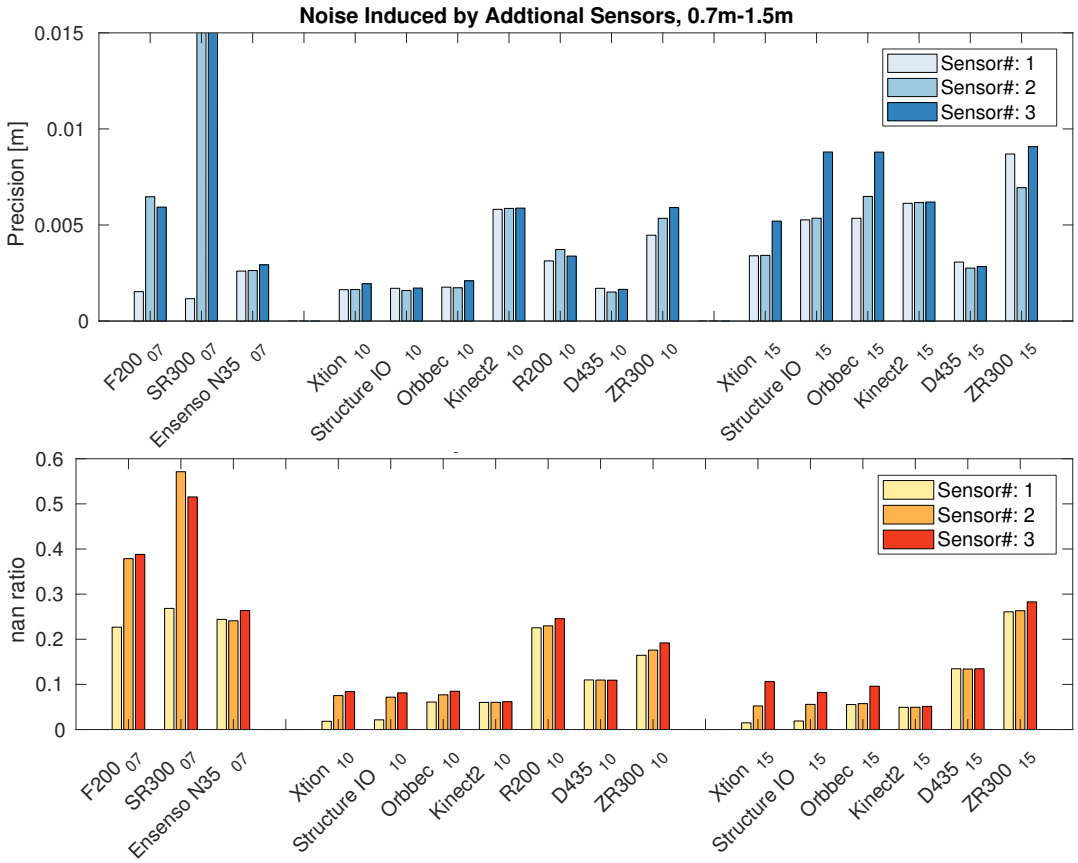


Fig. 6: Precision and nan ratio (lower is better) in multi sensor setup. Indices represent distance to the target: 07= $0.7m$, 10= $1m$,15= $1.5m$.

Fig. 7: Influence of additional sensors. SR300. Number of sensors observing the scene, from left to right: 1,2,3.

tered results for precision and bias.

For large distances $> 4m$, the tested ToF sensor (Kinect2) gathers the most reliable measurements, even under bright lighting conditions.

The Ensenso Active Stereo camera offers the lowest bias within its narrow range from $0.5 - 1m$. It satisfies applications requiring low biased measurements from a sensor that can be used out of the box.

The RealSense ZR300, R200, and D435 offer various parameters to adapt the sensor properties to the scene. During our experiments we used the factory presets for high accurate measurement without adaptation to the current lighting and material.

Experiments I and II (bias and precision) are easy to reproduce, we suggest users with new sensors to gather data in a similar way to i) benchmark their sensor against our results and ii) apply the introduced easy to use error model.

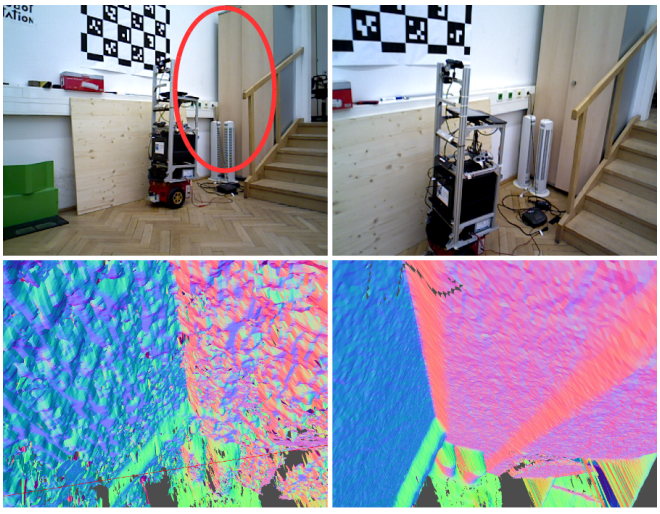


Fig. 8: Reconstruction method from [26]. The two rgb images show the same location observed from different distances. The bottom images show the reconstruction results of the highlighted region (red circle), without (left) and with (right) our error model.

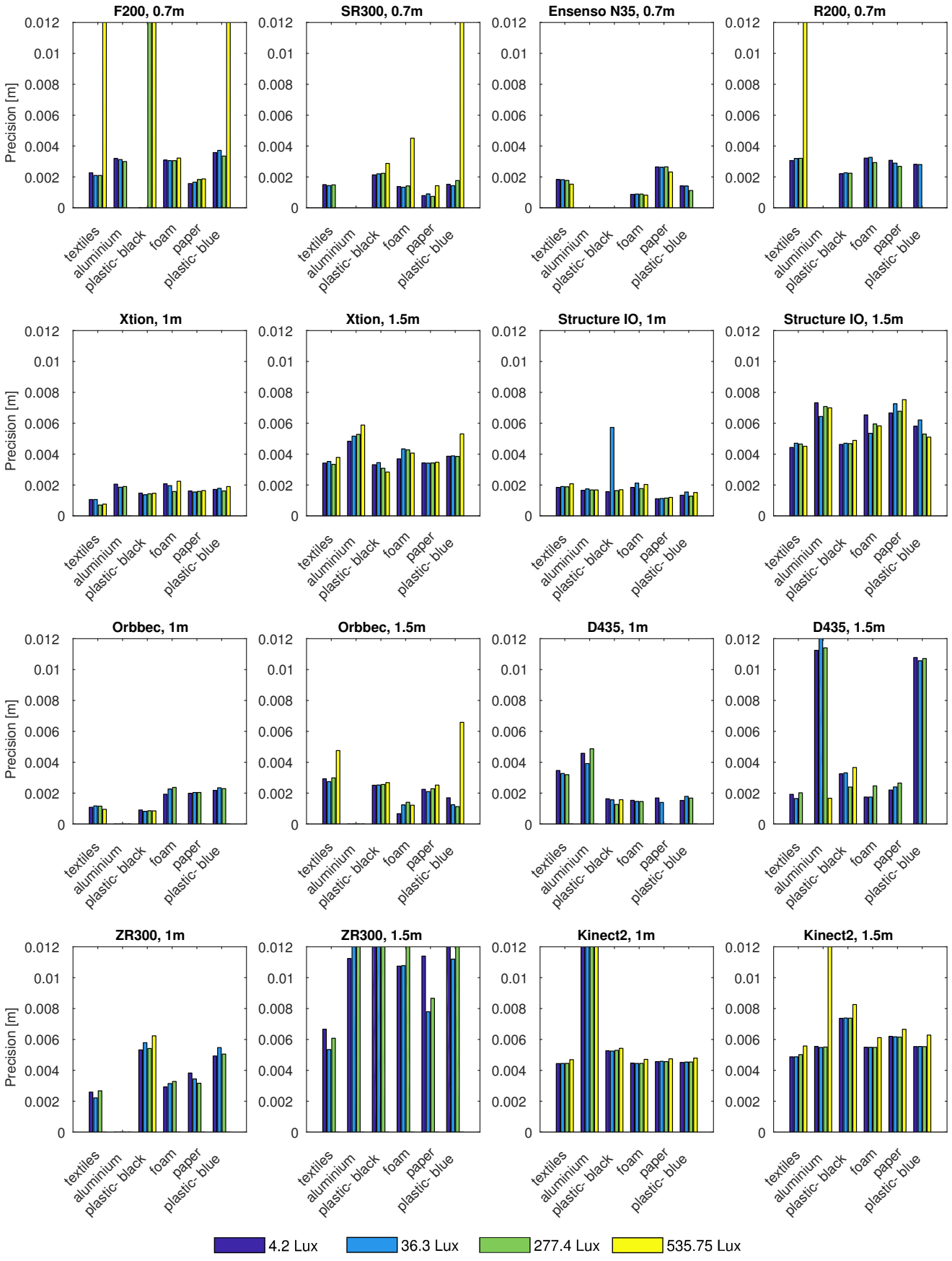


Fig. 9: Precision for different materials and lighting conditions. Lower is better. A precision of zero indicates that the sensor was not able to gather any depth information.

Future work should evaluate the different sensors under outdoor lighting conditions and add new sensors that continuously enter the market. Furthermore, the experiments regarding precision, bias, and lateral noise could be extended for different viewing angles. The setup may be extended by replacing standard drivers with more advanced methods. E.g., [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.

VIII. Acknowledgment

The authors would like to thank Farhoud Malekghasemi and Simon Schreiberhuber for their help with our experiments. The authors are also grateful to CogVis GmbH[2] 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 3d reconstruction and tracking," in *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*. IEEE, 2012, pp. 524–530.

[2] ASUS. (2017) Specifications Xtion Pro Live. Retrieved 2018-03-06. [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," Patent, Apr. 3, 2012, US Patent 8,150,142.

[4] Structure. (2017) Specifications Structure IO. Retrieved 2018-03-06. [Online]. Available: https://structure.io/embedded

[5] Orbbec. (2017) Specifications Orbbec Astra Pro. Retrieved 2018-03-06. [Online]. Available: https://orbbec3d.com/product-astra-pro/

[6] Microsoft. (2017) Specifications Microsoft Kinect. Retrieved 2018-03-06. [Online]. Available: https://developer.microsoft.com/en-us/windows/kinect/hardware

[7] Intel Corporation. (2018) Intel RealSense Camera D435. Retrieved 2018-03-06. [Online]. Available: https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-D400-Series-Datasheet.pdf

[8] ——. (2017) Intel RealSense Camera R300. Retrieved 2018-03-06. [Online]. Available: https://click.intel.com/media/ZR300-Product-Datasheet-Public-002.pdf

[9] ——. (2017) Intel RealSense Camera R200. Retrieved 2018-03-06. [Online]. Available: https://software.intel.com/sites/default/files/managed/d7/a9/realsense-camera-r200-product-datasheet.pdf

[10] ——. (2017) Intel RealSense Camera F200. Retrieved 2018-03-06. [Online]. Available: https://communities.intel.com/docs/DOC-24012

[11] ——. (2017) Intel RealSense Camera SR300. Retrieved 2018-03-06. [Online]. Available: https://software.intel.com/sites/default/files/managed/0c/ec/realsense-sr300-product-datasheet-rev-1-0.pdf

[12] Ensenso. (2017) Specifications N35 Series. Retrieved 2018-03-06. [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 Robotics and Automation Letters*, 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 *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 4465–4472.

[15] M. Nießner, M. Zollhöfer, S. Izadi, and M. Stamminger, "Real-time 3d reconstruction at scale using voxel hashing," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, p. 169, 2013.

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

[17] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, "International vocabulary of metrology–basic and general concepts and associated terms, 2008," *JCGM*, vol. 200, pp. 99–12, 2008.

[18] B. N. Taylor and C. E. Kuyatt, *Guidelines for evaluating and expressing the uncertainty of NIST measurement results*. US Department of Commerce, Technology Administration, National Institute of Standards and Technology Gaithersburg, MD, 1994.

[19] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," *IEEE transactions on cybernetics*, 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," *Electrical and Computer Engineering Technical Report ECE-TR-6*, 2012.

[21] J. Smisek, M. Jancosek, and T. Pajdla, "3D with Kinect," in *Consumer depth cameras for computer vision*. Springer, 2013, pp. 3–25.

[22] C. Pramerdorfer, "Depth data analysis for fall detection," 2013, TUW, Master Thesis.

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

[24] D. ISO, "5725-1: 1994," *Accuracy (trueness and precision) of measurement methods and results–part*, vol. 1.

[25] J. Canny, "A computational approach to edge detection," in *Readings in Computer Vision*. 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 *Proceedings of OAGM Workshop*, 2018.

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

[28] B. Zeisl and M. Pollefeys, "Structure-based auto-calibration of rgb-d sensors," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 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 *Proceedings of the IEEE Conference on 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)*. Springer International Publishing AG, 2016.

[2] CogVis GmbH: http://www.cogvis.at/