

EVALUATION OF INTEL REALSENSE D455 CAMERA DEPTH ESTIMATION FOR INDOOR SLAM APPLICATIONS

P. Hübner*, J. Hou, D. Iwaszczuk

Technical University of Darmstadt,
Dept. of Civil and Environmental Engineering Sciences,
Remote Sensing and Image Analysis, Darmstadt, Germany
(patrick.huebner, jiwei.hou, dorota.iwaszczuk)@tu-darmstadt.de

KEY WORDS: sensor evaluation, indoor mapping, indoor slam, rgbd

ABSTRACT:

The aim of this study is to propose evaluation methodology for the quality assessment of depth cameras for indoor mapping applications. Specifically, we evaluate the RGBD sensor Intel Realsense D455 w.r.t. measurement accuracy and noise while investigating their dependence on two central parameters characterizing the measurement scenario simultaneously: measurement distance and inclination of observed surfaces. The evaluation results are presented two-dimensionally as a function of both parameters. To this aim, two evaluation studies are conducted. First, a checkerboard pattern is used as reference object in a controlled setting. Secondly, actual indoor mapping data resulting from the deployment of the evaluated sensor in the context of an RGBD-SLAM application is used to conduct a similar evaluation in representative conditions for the intended usage scenario of indoor mapping. In this case, a TLS scan of the respective indoor environment is used as reference for evaluation.

1. INTRODUCTION

Low-cost 3D sensors such as RGBD cameras together with SLAM-based tracking provide the means for efficient and affordable mobile mapping systems with potential applications in fields like robotics and teleoperation (Stotko et al., 2019), documentation of cultural heritage (Teggi and Fassi, 2022), forest monitoring (Iwaszczuk et al., 2023) as well as the mapping (Hou et al., 2023) and automatic modeling (Weinmann et al., 2021; Schmidt et al., 2023) of indoor building environments.

Providing insights about the measurement quality that can be expected in different situation from a respective depth sensor is important as the quality of depth sensing potentially influences the mentioned downstream tasks. Thus, in this work, we target the topic of RGBD camera evaluation for indoor SLAM applications which aim at mapping and modeling indoor building environments.

Specifically, we focus on the joint consideration of measurement distance and inclination of observed surfaces towards the sensor as both directly influence the accuracy of depth measurements. While existing evaluation studies only vary distance in a perpendicular setting or additionally investigate the influence of inclination at a fixed distance, we propose to investigate measurement quality as a two-dimensional entity depending on both parameters simultaneously.

In this paper, we present an evaluation of the RGBD camera Intel Realsense D455. However, the proposed evaluation methodology is generic and can be applied to other depth cameras and potentially even to other kinds of active range sensors. The paper presents two

evaluation studies. One focuses on a checkerboard pattern as reference object and data recorded specifically for evaluation purposes in a controlled setting. The other focuses on using indoor mapping data itself for evaluation purposes by using a TLS scan of the respective indoor environment as reference.

After briefly discussing related work in Section 2, Section 3 presents details on the measurement setup and data analysis for both evaluation studies. Afterwards, the obtained results are presented in Section 4 and discussed in Section 5 before Section 6 concludes the paper.

2. RELATED WORK

To the best of our knowledge, there are no existing evaluation studies of depth cameras for indoor mapping applications or other usage scenarios where depth measurement accuracy is investigated in dependence of measurement distance and inclination angle of observed surfaces simultaneously. Oftentimes, only the measurement distance is considered as a parameter representing the measurement scenario while keeping a fixed, perpendicular angle of observation towards a reference plane (Khoshelham and Oude Elberink, 2012; Fürsattel et al., 2016; Carfagni et al., 2017). If inclination is considered at all, it is evaluated separately with fixed measurement distance (Hübner et al., 2020). In terrestrial laser scanning, the influence of surface inclination on measurement quality is more commonly considered (Soudarissanane et al., 2011; Schmitz et al., 2019).

Surface inclination is sometimes considered implicitly by evaluating against three-dimensional reference objects with varying surface orientations (Khoshelham and Oude Elberink, 2012; Servi et al., 2021). However, in this case, the resulting accuracies are represented as functions of

* Corresponding author

image or 3D coordinates and not in distance/inclination space as we propose to do for sensor characterization. A similar situation is given for evaluation studies that focus on complete indoor mapping systems instead of just the depth sensing part (Lehtola et al., 2017; Assali et al., 2019; Hübner et al., 2019). Here, the resulting point clouds are evaluated in 3D against suitable reference data without, however, considering the measurement conditions of the individual points. More specifically, it is not investigated, under which distance and angle of inclination a respective point has been measured and how this affects its measurement accuracy.

3. EVALUATION METHODOLOGY

This section presents the methodology for two evaluation studies aiming at characterizing the quality of the depth camera Intel Realsense D455 in terms of noise and accuracy over the whole measurement range w.r.t. the parameters of distance and angle of inclination of observed surfaces. First, Section 3.1 presents the methodology for an evaluation study in a controlled setting using a checkerboard pattern as reference object. Afterwards, Section 3.2 presents methodology for doing a similar evaluation study directly on indoor mapping data acquired by using the depth camera while relying on a TLS point cloud of the respective indoor environment as reference dataset.

In both experiments, the Intel Realsense D455 device was used after having been in use for over an hour to avoid heating effects. Furthermore, the calibration routine implemented in the Intel Realsense Viewer was applied in order to calibrate the device.

3.1 Evaluation Against A Checkerboard

This section presents the applied evaluation methodology with a checkerboard pattern as reference object. First, Section 3.1.1 presents details on the evaluation setup. Afterwards, Section 3.1.2 presents the methodology applied for evaluating the recorded data.

3.1.1 Measurement Setup In this experiment, a checkerboard pattern of 5×7 squares of 9.5 cm size printed on a rigid cardboard slide affixed to a wooden board was placed against a wall standing on a table. This reference object was subsequently observed with the Intel Realsense D455 device from different standpoints while color and depth frames of 1280×720 pixels were recorded to a rosbag using the Intel Realsense Viewer.

The selection of standpoints aims at covering a wide range of distances and angles of observation w.r.t. the reference plane defined by the checkerboard pattern. The measurement process repeatedly started ca. 1 m in front of the checkerboard pattern and subsequently increased the distance in steps of ca. 40 cm until reaching the end of the measurement range where most pixels of the depth frames returned no value. In this manner, first, a trajectory was recorded approximately perpendicular to the checkerboard plane. Subsequently, this process was repeated with increasing inclinations towards the checkerboard plane. At each standpoint, depth and color frames were recorded in a static setting for ca. 1.5 minutes each.

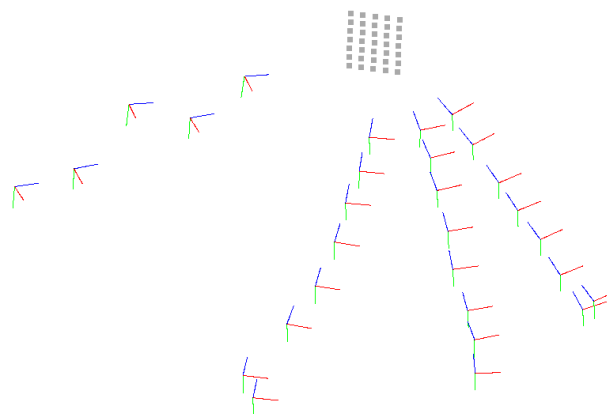


Figure 1. The poses (x, y and z axis visualized in red, green, blue respectively; axis length 20 cm) of the Realsense D455 color camera relative to the coordinate frame defined by the checkerboard points depicted in grey. The depicted poses are averaged over all 150 images per standpoint used for evaluation.

3.1.2 Data Analysis As input data for the evaluation, a fixed set of 150 frames each was considered per standpoint. In case more frames were recorded for a standpoint, these were discarded in order to keep the number of frames comparable. The color and depth frames stored independently in the rosbag files were temporally aligned and spatially aligned by using the utility methods provided by the Intel Realsense SDK. Thus, pixel-aligned RGBD frames were used for evaluation.

In order to determine reference poses per standpoint of the Intel Realsense D455 device relative to the reference pattern, the checkerboard detection from OpenCV and the EPnP algorithm (Lepetit et al., 2009) were applied per color frame. The resulting per-frame poses were subsequently averaged over all 150 frames to obtain a reference pose per standpoint. These poses are depicted in Figure 1. The variation in the resulting reference distances to the checkerboard plane over all frames per standpoint are reported as part of the evaluation results.

Besides for obtaining reference poses and thus reference values for distance and inclination of the checkerboard plane, the checkerboard detection also serves for restricting the evaluation to those pixels and their respective depth values, that depict the checkerboard plane. In this way, a planar measurement target could be guaranteed and no manual masking of pixels not belonging to the reference plane was necessary.

For evaluating noise, i.e. precision of the depth camera, a plane was fitted into the depth points covering the checkerboard plane via Singular Value Decomposition (Brown, 1976) for each depth frame. Orthogonal distances of each depth point to this plane were evaluated for each frame with averaged values over all frames per standpoint reported in the evaluation results. Besides noise, characterizing the variation of measurements from the best-fitted plane, the accuracy of the sensor was evaluated by comparing the points against a reference plane determined by the checkerboard pattern and its known dimensions. In Figure 2, depth points and the reference points defining the checkerboard pattern are depicted for an exemplary frame

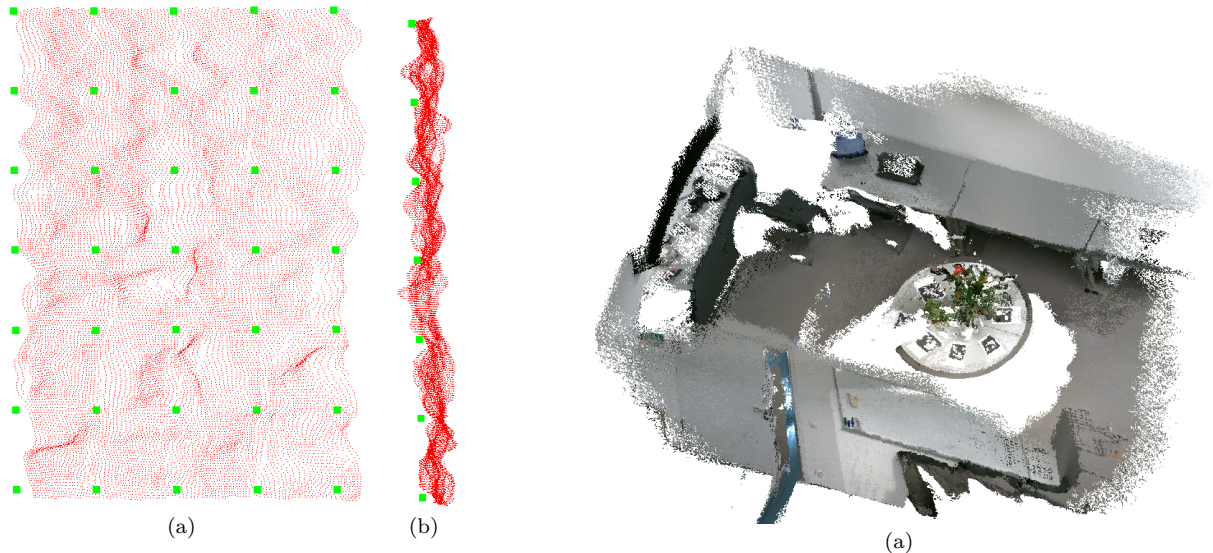


Figure 2. Visualization of checkerboard points (green) and Realsense D455 depth points (red) for a standpoint at ca. 1.7 m distance and an inclination of ca. 42.5°. (a) front view (b) side view.

from a standpoint at ca. 1.7 m distance and an inclination of ca. 42.5°. The resulting values for noise and accuracy were averaged over all checkerboard pixels per frame and over all frames per standpoint.

The presented evaluation procedure was applied separately for the pixels depicting the black and white squares of the checkerboard pattern in order to investigate the influence of the color on measurement noise and accuracy.

3.2 Evaluation Against A Reference Scan

In a second experiment, actual indoor mapping data acquired with the Intel Realsense D455 device is used for evaluation instead of an artificial reference object like the checkerboard. In this case, a point cloud of the same indoor environment acquired by terrestrial laser scanning (TLS) is used as reference. As before in Section 3.1, first, Section 3.2.1 presents details on the measurement setup before Section 3.2.2 elaborates on evaluation methodology.

3.2.1 Measurement Setup In this experiment, a conference room inside a university building depicted in Figure 3 was chosen as object of study. An RGBD sequence with 1334 frames was recorded in slow walking speed with 30 frames per second at 640×480 pixel resolution for a trajectory around the round table in the center of the room was recorded with the device pointing in walking direction.

As reference dataset, a point cloud was acquired with a terrestrial laser scanner (Zoller & Fröhlich Imager 5016). In total, 4 standpoints around the round table in the center of the room were scanned and the respective point clouds were fused via the integrated on-site real-time registration of the device to the point cloud depicted in Figure 3(b).

3.2.2 Data Analysis For the recorded RGBD sequence, poses were estimated via ORB-SLAM3 RGBD-SLAM (Campos et al., 2021). The obtained poses allow

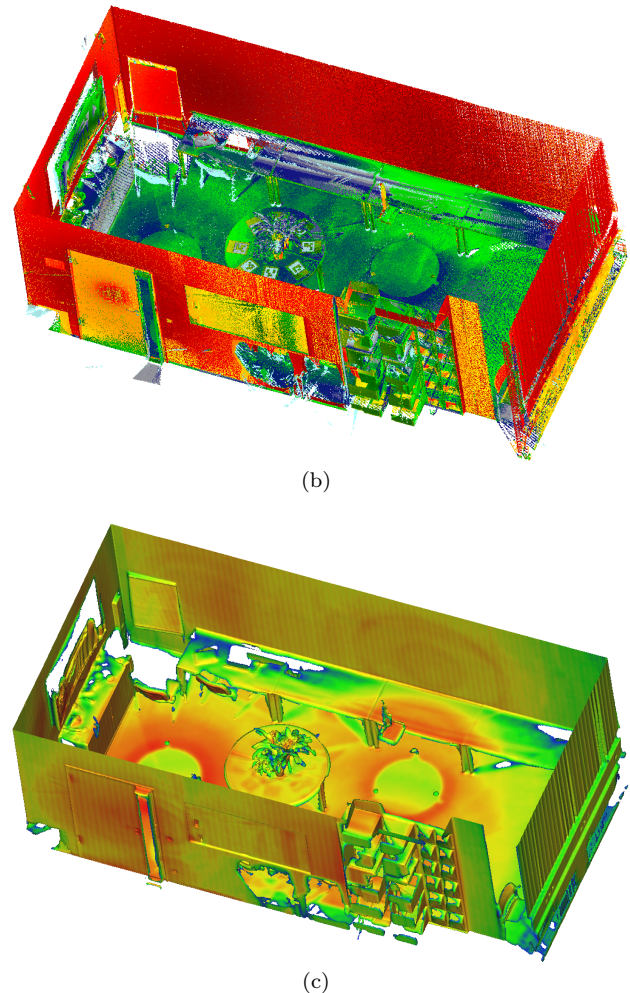


Figure 3. (a) Subset of the Intel Realsense D455 pointcloud used for registration (b) TLS reference point cloud (the color scale represents intensity and is optimized for contrast) (c) triangle mesh generated from the TLS point cloud after applying a threshold on point density for removing mesh artefacts resulting from the poisson algorithm (the color scale represents point density)

to project the depth images into 3D space and thus to create an indoor mapping point cloud of the room.

In order to evaluate the single depth frames against the TLS reference point cloud, the frame poses need to be transformed to the coordinate system of the TLS point cloud. To ensure best possible registration quality, only pixels with a depth between 0.5 and 2.0 m were selected for creating the point cloud depicted in Figure 3(a). Due to this, the right part of the room is missing in the depicted point cloud as distances are here all larger 2 m. This point cloud is however only used for registration. In the subsequent evaluation, depth pixels of all frames and with values in the whole measurement range of up to 6 m according to device specifications are used.

The registration was performed by manually selecting pass points in Cloud Compare and subsequent fine registration with ICP (Besl and McKay, 1992). The resulting registration with an estimated scale factor of 0.98288 results in mean cloud to cloud distance of 2.7 cm.

In order to evaluate the posed depth camera frames against the reference point cloud, projected depth points need to be assigned a corresponding point from the reference point cloud. As searching for the nearest point in 3D space was found to be too imprecise for large measurement distances strongly affected by noise, ray casting needs to be applied. In order to efficiently implement this, a triangle mesh was created from the reference point cloud by applying Poisson surface reconstruction (Kazhdan et al., 2006) with subsequent thresholding on point density to remove artifacts. The resulting triangle mesh is depicted in Figure 3(c).

For each pixel in the depth images, ray tracing was performed against the triangle mesh based on the respective frame pose and the calibrated inner orientation of the camera. If a ray intersects the triangle mesh, the reference distance can be determined and the inclination can be derived from the normal vector of the intersecting triangle. The reference distances were compared to the respective depth measurement of the Intel Realsense D455 device to obtain an accuracy value for the respective reference distance and inclination. The obtained accuracy values over all frames were binned in a distance/inclination grid with $5\text{ mm} \times 0.1^\circ$ resolution while keeping track on an iterative statistics with count, mean value and variance per grid cell.

4. RESULTS

In this section, the results obtained by the evaluation methods proposed in Section 3 are presented. First, Section 4.1 presents the results of the evaluation study against a checkerboard pattern as reference object. Afterwards, Section 4.2 presents results for the evaluation against a TLS reference scan.

4.1 Checkerboard

Figure 4 visualizes the standard deviations of reference distances per standpoint over all 150 frames respectively corresponding to the averaged poses per standpoint depicted in Figure 1.

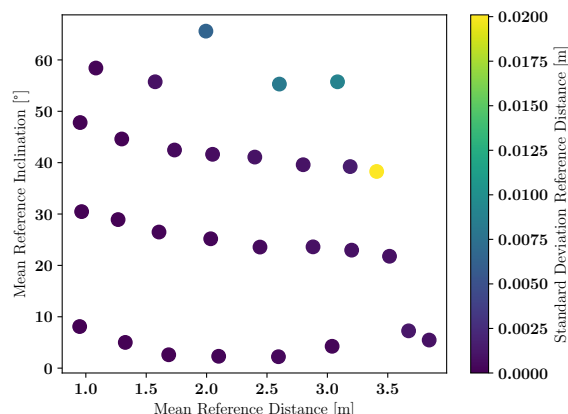


Figure 4. Standard deviation of the reference distances over all images per standpoint estimated from the observed checkerboard pattern. The noise in the reference distance results from the pose estimation from checkerboard points which becomes unstable at steep angles of observation.

Figure 5 exemplarily depicts pixel-wise noise and accuracy values for an example frame of a standpoint at ca. 1.7 m distance and an inclination of ca. 42.5° corresponding to the points depicted in Figure 2. Averaged values per standpoint are visualized in Figure 6 for noise and Figure 7 respectively.

The presented values refer to an evaluation performed over all pixels of the checkerboard pattern, irrespective of color. Separate evaluations for the black and white squares of the checkerboard pattern have been performed. However, all differences in resulting mean values have been found to be below their respective standard deviations and can thus be considered as not significant.

4.2 Reference Scan

Exemplary results of the evaluation study using indoor mapping data and a TLS point cloud as reference are depicted in Figure 8 for a single frame of the sequence. Figure 9 shows the accuracy results for the same frame as distance/inclination grid visualizing sample count, mean accuracy and standard deviation per grid cell. The aggregated results over all pixels of the whole image sequence are presented in Figure 10. In order to allow focusing on meaningful areas, the visualizations of accuracy and its standard deviation per cell are thresholded by sample count.

5. DISCUSSION

The evaluation results presented in Section 4 are discussed in the following. First, Section 4.1 discusses the results of the evaluation study against a checkerboard pattern before Section 4.2 discusses the results of the evaluation against a reference scan.

5.1 Checkerboard

For most standpoints, the noise in the reference distances is well below 5 mm. Only in the higher regions of the distance/inclination space, heightened standard deviations

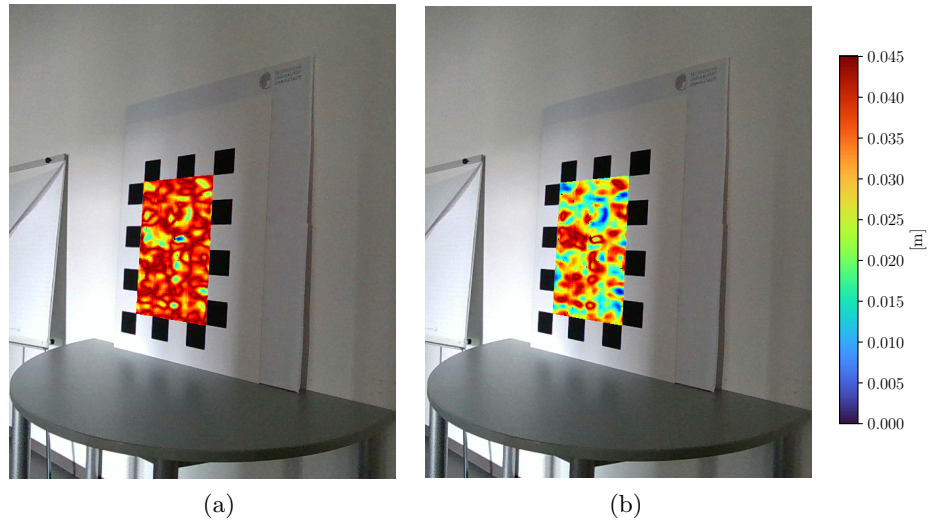


Figure 5. Visualization of (a) noise and (b) accuracy of an exemplary frame captured at ca. 1.7 m distance and an inclination of ca. 42.5° (cropped) corresponding to the points depicted in Figure 2.

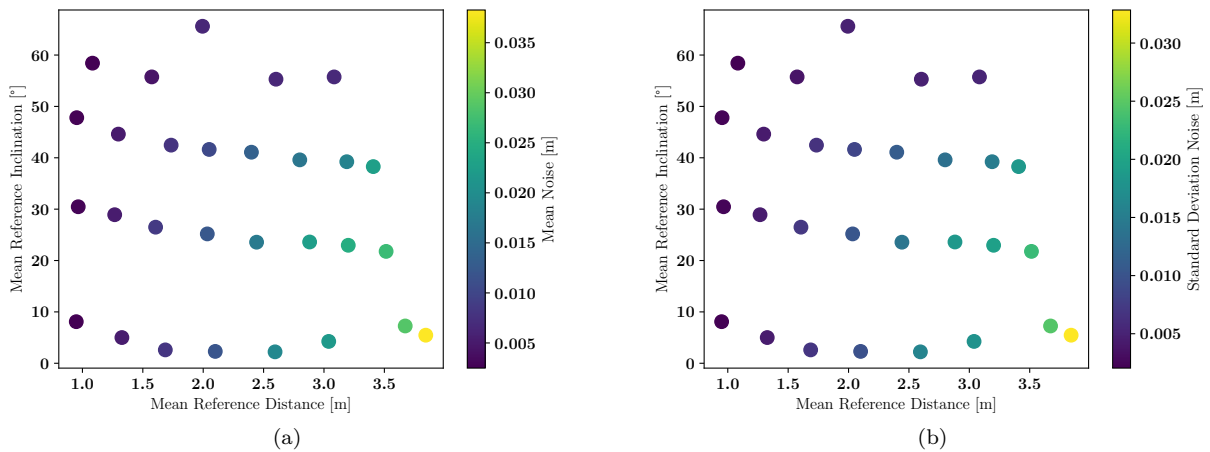


Figure 6. Mean (a) and standard deviation (b) of noise per standpoint evaluated against a plane fitted to the measured depth points on the checkerboard.

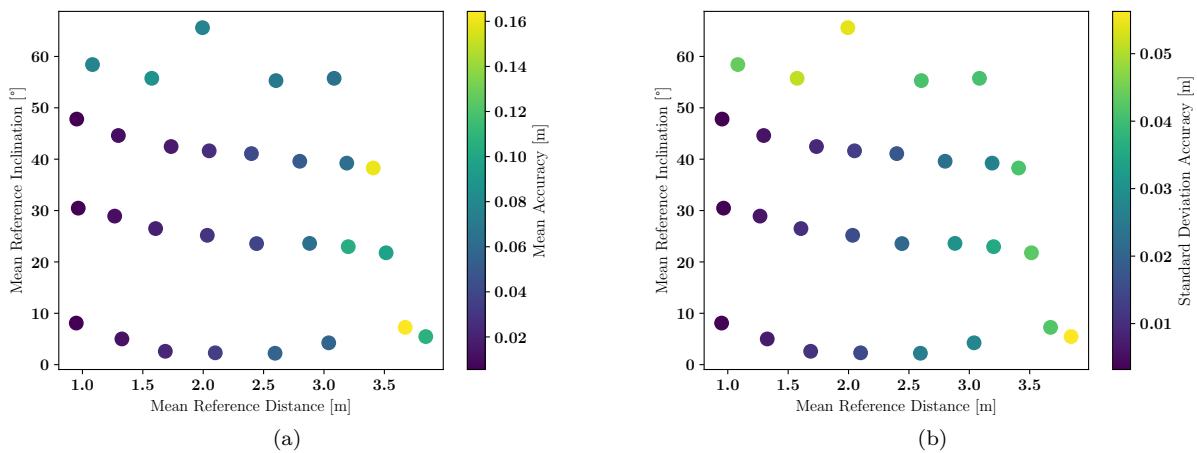


Figure 7. Mean (a) and standard deviation (b) of accuracy per standpoint evaluated against the checkerboard plane.

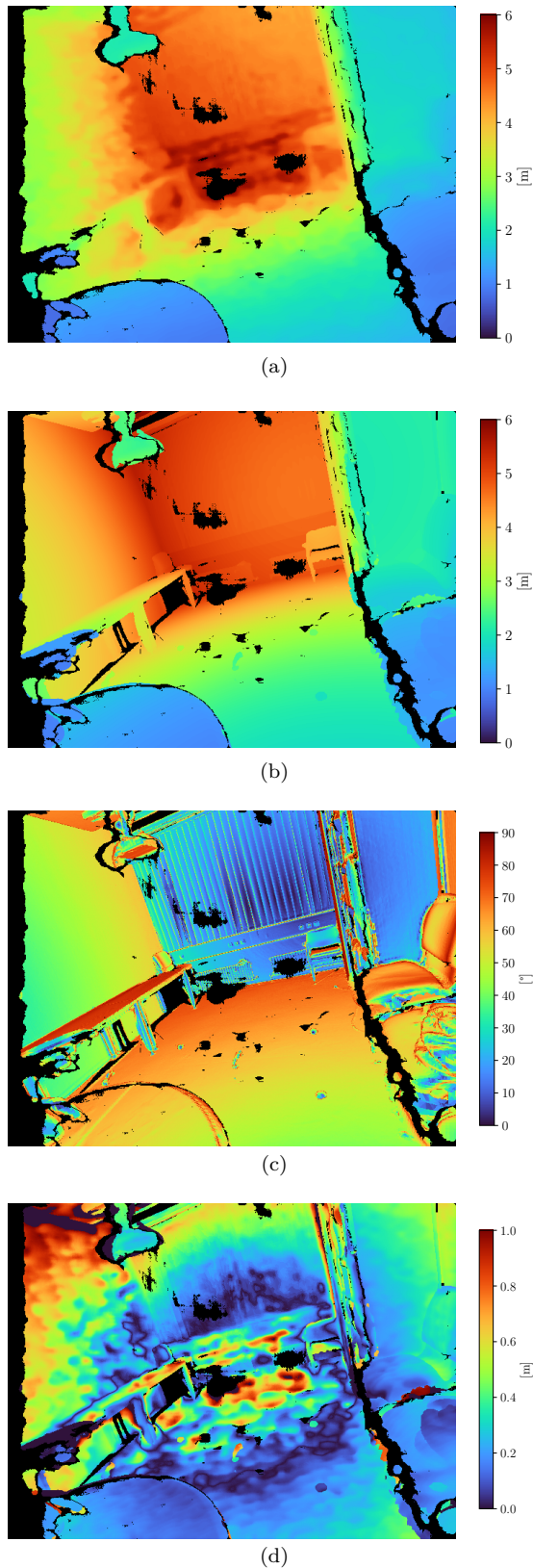


Figure 8. Results for an exemplary frame. (a) Intel Realsense D455 depth (b) reference depth from the triangle mesh (c) surface angle from the triangle mesh (d) depth accuracy. In (a), black pixels do not hold values or are above the nominal working distance of 6 m of the device. In (b) to (d), black pixels can additionally result from not finding a hit with the ground truth mesh during ray casting.

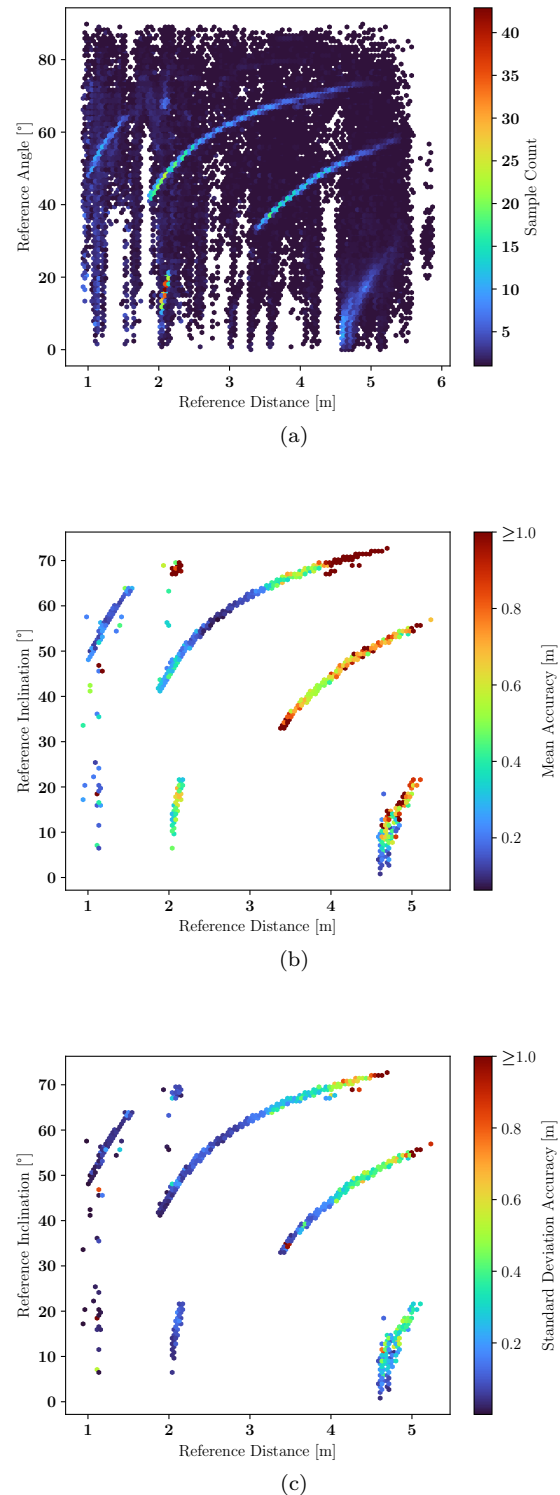
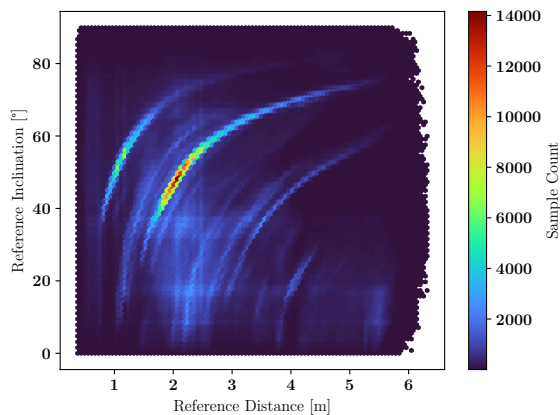
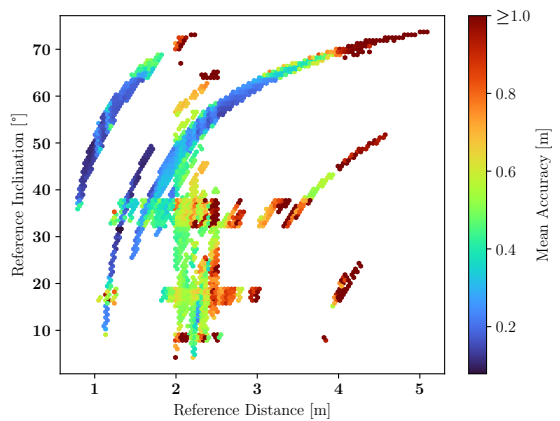


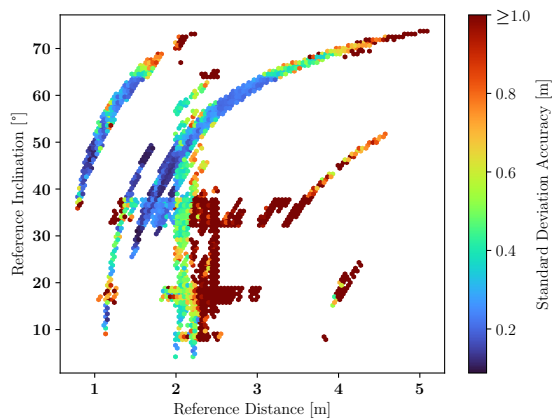
Figure 9. Visualization of (a) sample count, (b) mean accuracy and (c) standard deviation of Intel Realsense D455 depth measurements of all pixels of the exemplary frame depicted in Figure 8(a) evaluated against a reference scan. In order to allow focusing on meaningful areas, only cells with a sample count of at least 10 are depicted in (b) and (c).



(a)



(b)



(c)

Figure 10. Visualization of (a) sample count, (b) mean accuracy and (c) standard deviation of Intel Realsense D455 depth measurements of all pixels of all 1334 images evaluated against a reference scan. In order to allow focusing on meaningful areas, only cells with a sample count of at least 2000 are depicted in (b) and (c).

of up to 2 cm can be observed. These result from the pose determination from the checkerboard pattern becoming unstable at steep angles of observation when the distance is large at the same time. This uncertainty in pose determination needs to be considered when interpreting the results for noise and accuracy.

The results presented in Figure 6 and 7 show that the deviations from the reference planes are about three to four times larger than the noise. In both cases, high mean values for noise and accuracy correspond with a high variance among the results of the single frames of a standpoint. Furthermore, it can be observed, that noise mainly seems to increase with distance while there is no apparent increase in noise with inclination. Accuracy on the other hand seems to worsen with higher distances as well as with inclination.

5.2 Reference Scan

The visualizations of the samples per cell in the distance /inclination grid, for the single frame example in Figure 9(a) as well as in the aggregated grid over all frames in Figure 10(a) show distinctive lines. It might be assumed that these correspond to the main room surfaces which often seem to be depicted in approximately similar viewing conditions over the course of the circular trajectory through the center part of the room.

The visualization for accuracy in Figure 9(b) and noise (as which the standard deviation of the accuracy per cell can be interpreted) in Figure 9(c) for the exemplary single frame shows comparable characteristics to the results of the checkerboard experiment discussed in Section 5.1. While noise mainly increases with high distances, accuracy shows an additional sharp degradation in high inclinations.

The results for the aggregated grid in Figure 10(b) and 10(c) are harder to interpret. Here, high inclinations seem to cause a breakdown in accuracy as well as in noise. Furthermore, there is a further apparent deterioration of quality (again in noise as well as in accuracy) for low inclinations starting with moderate distances from about 2.5 m and higher.

Generally, it should be further investigated, how much the shown results actually represent sensor quality and to what degree they are effected by errors in pose determination, in the registration between the depth frames and the reference point cloud and in the meshing of the point cloud. The depth images in Figure 8(a) and 8(b) mostly suggest a quite good quality of registration and reference mesh which can be visually assessed on the masked out black regions from Figure 8(a) which are also visible in Figure 8(b). There are deviations in the vicinity of the projector mounted on the ceiling. The contour of the table and the furniture on the right-hand side of the images seem to suggest a good overlay. Further investigations on this matter remain subject of future research.

6. CONCLUSION

In this paper, we present methodology for evaluating RGBD cameras regarding depth measurement noise and

accuracy w.r.t. the task of mapping indoor environments. Indoor mapping scenarios typically involve measurement conditions comprising a wide range of distances and inclinations as demonstrated in Figure 10(a). Thus, we propose to regard sensor quality as a two-dimensional field depending on the parameters distance and inclination instead of investigating both parameters separately. The proposed methodology was applied to investigating the Intel Realsense D455 depth camera. The presented results show that while evaluation in a controlled setting hints at good quality over a wide range of distance and inclination conditions, it is still partly an open research question, how to actually assess indoor mapping quality in real-usage conditions in a detailed pixel-wise manner.

References

- Assali, M., Pipelidis, G., Podolskiy, V., Iwaszczuk, D., Heinen, L., Gerndt, M., 2019. Quantifying the Quality of Indoor Maps. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W13, 739-745.
- Besl, P., McKay, N. D., 1992. A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 239-256.
- Brown, C. M., 1976. Principal Axes and Best-Fit Planes, with Applications. Technical report, University of Rochester. Computer Science Department.
- Campos, C., Elvira, R., Gómez Rodríguez, J. J., Montiel, J. M. M., Tardós, J. D., 2021. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM. *IEEE Transactions on Robotics*, 37(6), 1874-1890.
- Carfagni, M., Furferi, R., Governi, L., Servi, M., Uccheddu, F., Volpe, Y., 2017. On the Performance of the Intel SR300 Depth Camera: Metrological and Critical Characterization. *IEEE Sensors Journal*, 17(14), 4508-4519.
- Fürsattel, P., Placht, S., Balda, M., Schaller, C., Hofmann, H., Maier, A., Riess, C., 2016. A Comparative Error Analysis of Current Time-of-Flight Sensors. *IEEE Transactions on Computational Imaging*, 2(1), 27-41.
- Hou, J., Goebel, M., Hübner, P., Iwaszczuk, D., 2023. Octree-Based Approach for Real-Time 3D Indoor Mapping Using RGB-D Video Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1/W1-2023, 183-190.
- Hübner, P., Clintworth, K., Liu, Q., Weinmann, M., Wursthorn, S., 2020. Evaluation of HoloLens Tracking and Depth Sensing for Indoor Mapping Applications. *Sensors*, 20(4), 1021:1-24.
- Hübner, P., Landgraf, S., Weinmann, M., Wursthorn, S., 2019. Evaluation of the Microsoft HoloLens for the Mapping of Indoor Building Environments. *Dreiländertagung der DGPF, der OVG und der SGPF in Wien, Österreich – Publikationen der DGPF*, 28, 44–53.
- Iwaszczuk, D., Goebel, M., Du, Y., Schmidt, J., Weinmann, M., 2023. Potential of Mobile Mapping to Create Digital Twins of Forests. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1/W1-2023, 199-206.
- Kazhdan, M., Bolitho, M., Hoppe, H., 2006. Poisson Surface Reconstruction. *Eurographics Symposium on Geometry Processing (SGP)*, 61–70.
- Khoshelham, K., Oude Elberink, S., 2012. Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. *Sensors*, 12(2), 1437-1454.
- Lehtola, V. V., Kaartinen, H., Nüchter, A., Kaijaluoto, R., Kukko, A., Litkey, P., Honkavaara, E., Rosnell, T., Vaaja, M. T., Virtanen, J.-P., Kurkela, M., El Issaoui, A., Zhu, L., Jaakkola, A., Hyyppä, J., 2017. Comparison of the Selected State-Of-The-Art 3D Indoor Scanning and Point Cloud Generation Methods. *Remote Sensing*, 9(8), 796:1-26.
- Lepetit, V., Moreno-Noguer, F., Fua, P., 2009. EPnP: An Accurate O(n) Solution to the PnP Problem. *International Journal of Computer Vision*, 81(2), 155–166.
- Schmidt, J., Volland, V., Iwaszczuk, D., Eichhorn, A., 2023. Detection of Hidden Edges and Corners in SLAM-Based Indoor Point Clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1/W1-2023, 443-449.
- Schmitz, B., Holst, C., Medic, T., Lichti, D. D., Kuhlmann, H., 2019. How to Efficiently Determine the Range Precision of 3D Terrestrial Laser Scanners. *Sensors*, 19(6), 1466.
- Servi, M., Mussi, E., Profili, A., Furferi, R., Volpe, Y., Governi, L., Buonamici, F., 2021. Metrological Characterization and Comparison of D415, D455, L515 RealSense Devices in the Close Range. *Sensors*, 21(22), 7770:1-19.
- Soudarissanane, S., Lindenbergh, R., Menenti, M., Teunissen, P., 2011. Scanning Geometry: Influencing Factor on the Quality of Terrestrial Laser Scanning Points. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(4), 389-399.
- Stotko, P., Krumpfen, S., Schwarz, M., Lenz, C., Behnke, S., Klein, R., Weinmann, M., 2019. A VR System for Immersive Teleoperation and Live Exploration with a Mobile Robot. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3630–3637.
- Teruggi, S., Fassi, F., 2022. HoloLens 2 Spatial Mapping Capabilities in Vast Monumental Heritage Environments. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-2/W1-2022, 489-496.
- Weinmann, M., Wursthorn, S., Weinmann, M., Hübner, P., 2021. Efficient 3D Mapping and Modelling of Indoor Scenes with the Microsoft HoloLens: A Survey. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 89, 319-333.