11 Spatially Resolved Depth Sensing

It is common practice for stereo sensors to contain two cameras, C_1 and C_2, of known relative spatial location. In the event of parallel optical axes, this location is defined as the baseline B. To compute depth, pixels in image of C_1 are matched to pixels of C_2, forming correspondence pairs. For every correspondence pair, the depth d is computed from the disparity D, which represents the difference between their pixel positions [Giancola et al. 2018]:

$$d = f \frac{B}{D} \; . \tag{11}$$

Spatial Resolution. The depth resolution δ_z of spatially resolved sensors is limited by the disparity resolution ΔD [Zanuttigh et al. 2016]:

$$\delta_z = \frac{z^2}{Bf} \Delta D \; . \tag{12}$$

We denote the ground-truth depth as z and the focal length as f. The disparity resolution is dependent on δ_x and δ_y. For camera-based systems, δ_x and δ_y are typically expressed through the optical transfer function (OTF) [Williams and Becklund 2002].

12 Time-resolved Sensors (Time-of-Flight)

Time-of-Flight sensors can be roughly categorized into direct Time-of-Flight (dToF) and indirect Time-of-Flight (iToF) depth sensing methods. DToF sensors transmit a signal pulse and directly measure the time it takes for the pulse to return. Due to their high cost, however, they are less commonly used in close-range applications. More cost-efficient than dTof are continuous wave (CW) signal modulations that measure time indirectly (iToF) based on the phase shift $\Delta\varphi$ between the transmitted and received signal [Zanuttigh et al. 2016]. The general form of a continuous sinusoidal carrier signal s_c can be described by two equal formulas of traveling time t and traveling distance ρ, respectively:

$$s_c(t) = A \cdot \cos(2\pi t f + \phi_c) \tag{13}$$

$$= A \cdot \cos(2\pi \frac{\rho}{c} f + \phi_c) = \widehat{s_c}(\rho) \; . \tag{14}$$

A and f are the known signal amplitude and frequency, respectively and c is the speed of light, φ is the phase and ϕ_c is a constant phase offset. As a transmitted signal $s_t = s_c(t_1) = \widehat{s_c}(\rho_1)$ of known phase and amplitude reflects at a target, the received signal $s_r = s_c(t_2) = \widehat{s_c}(\rho_2)$ has a relative traveling distance of $2 \cdot \Delta\rho = (\rho_2 - \rho_1)$ between the transmitter and receiver. The range, $\Delta\rho$, is related to the relative phase shift $\Delta\varphi$ [Zanuttigh et al. 2016] by:

$$\Delta\rho = c \frac{\Delta\varphi}{4\pi f} \; . \tag{15}$$

The general assumption of dToF sensors is that a signal directly reflects at the first target and therefore the range equals half of the traveling distance. The depth resolution of a ToF sensor is specific to the utilized wavelength and spatial arrangement of transmitters and receivers.

NIR AMCW Time-of-Flight. AMCW ToF algorithms usually operate on the SIMO principle, as they do not require as expensive sensor apertures as imaging radars, and often have more receivers and transmitters than can be effectively managed computationally in MIMO depth estimation algorithms [Zanuttigh et al. 2016]. The range resolution of a NIR AMCW ToF sensor can be expressed as [Lopez Paredes et al. 2023]:

$$\delta_z = \frac{c}{f_m} \sqrt{\frac{P_l + P_a}{P_l} \cdot \frac{I}{k_o q_e \rho \Delta t}} \; . \tag{16}$$

Environment-specific parameters are the power of ambient light P_a, and the reflectivity of the target ρ. Hardware-specific parameters are the modulation frequency f_m, the power of the illumination unit P_l, the total illumination area I, the quantum efficiency q_e, the integration time Δt, and a constant parameter for the optical system, k_o. Due to unknown hardware-specific parameters, we were unable to determine the exact range resolution for NIR AMCW ToF (Azure Kinect) in Table 1 of the main paper. We refer to [Lopez Paredes et al. 2023] for an experimental approach of determining the effective range and lateral resolution.

MIMO FSCW Time-of-Flight. The spatial resolution of a square-shaped MIMO FSCW imaging sensor can be expressed as [Ahmed 2021]:

$$\delta_{x,y} = \frac{c}{4f_{max}} \cdot \sqrt{4\left(\frac{z}{L}\right)^2 + 1} \tag{17}$$

$$\delta_z = \frac{0.5 \cdot c}{\Delta f + \left(1 - \frac{1}{\sqrt{1 + 0.5(L/z)^2}}\right) \cdot f_{min}} \; . \tag{18}$$

We denote the size of the square aperture as L.

13 Sensor Parameters and Settings

The sensor settings in Table 1 of the main paper are chosen with respect to a trade-off between fair sensor comparability and practical applicability. We uniformly list the frame rate computed from the time takes to *capture* the relevant data of one depth frame. Note that this may not necessarily include the computation of depth. For instance, the QAR50 has a capture rate of ≈ 70 fps while the back-projection algorithm has an average computing time of 78 s such that the overall frame rate is below 1 fps. Furthermore, we manually adjusted each optical sensor's exposure time, if possible, to ensure similar lighting conditions. In summary, we selected the sensor settings that optimize quality while, when feasible, maintaining a comparable frame rate to that of the other sensors. Additionally, we adhered to the manufacturer's recommendations for optimal practical use in interactive applications.

13.1 Radar Field of View

Optical sensors typically model the field of view using a perspective camera model. In theory, MIMO radars can also be viewed as an array of small cameras such that the antenna aperture acts as a unified perspective camera, with its field of view defined by the union of all individual antenna frustums. In practice, modeling the complex antenna radiation pattern as a conventional camera

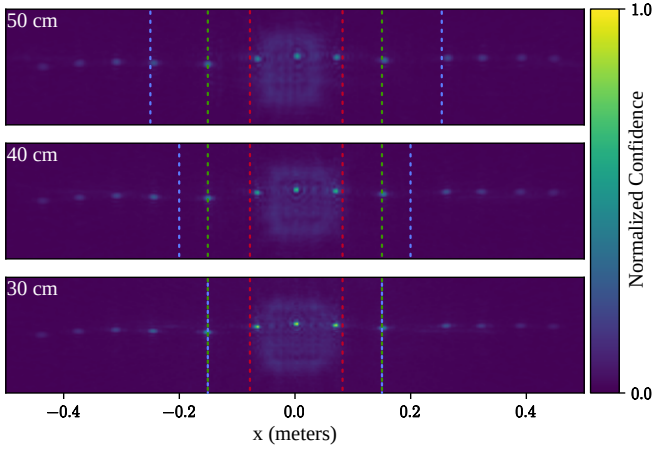Fig. 9. Styrofoam board with mounted metal spheres of ⌀1 cm.

Fig. 10. Confidence map of multiple point scatterers, displaced along the x-axis of the antenna aperture for the three object-to-sensor distances of MAROON. The confidence values are normalized across the three reconstructions. The vertical, dashed lines mark the horizontal extent of the antenna aperture, reconstruction volume, and approximated 53 × 53° perspective camera frustum.

frustum is a crude approximation, as the extents of the visible area are not as straightforward to define as for optical sensors.

To demonstrate this, we conducted an experiment where we mounted several 1 cm diameter metal spheres on a styrofoam board at a fixed horizontal distance around the aperture origin, as shown in Figure 9. The aim of this experiment is to explore the maximum visible area by measuring the signal response of each pair of spheres placed on opposite sides of the origin. The signal response for each metal sphere is illustrated in Figure 10, as part of the confidence value after spatially resolving the raw signal using backprojection.

For point targets with uniform, view-independent scattering properties, the imaging radar's visible area encompasses all the mounted metal spheres, covering an horizontal area of approximately 90 cm (and potentially even further).

In contrast, targets with extended surfaces and non-uniform scattering properties are reconstructed within a more limited area that roughly corresponds to the size of the antenna aperture (marked in red). Here, this is observed for the styrofoam board, which reflects

a minimal amount of the emitted signal and is typically considered nearly invisible. In this scenario, where only empty space is reconstructed alongside point targets, the signal response of the styrofoam board behaves similarly to other planar surface targets in MAROON.

In summary, the visible area of a MIMO radar has similarities with a continuous Gaussian function centered at the aperture origin. To determine the effective visible area, we compute the *full width at half maximum* (FWHM); here, it represents the horizontal extent of the reconstruction area where confidence values exceed 50 % of the maximum. For the three object-to-sensor distances of 30 cm, 40 cm, and 50 cm, this extent is approximately between [−0.15, 0.15] meters, aligning closely with the green-marked reconstruction volume used for evaluation. The corresponding fields of view of 53°, 41°, and 33° differ significantly across these distances, making it challenging to find a unified perspective camera frustum. A suitable perspective camera frustum would also need to fully encompass the 13.8 × 13.8 cm aperture at a distance of 0 cm, i.e., the aperture origin. Contrary to camera-based systems, where the spatial origin usually lies within the sensor extents, however, this frustum would yield an approximate field of view of 65°, with the camera origin located ≈ 10.8 cm behind the aperture. These observations suggest that an orthographic camera model, as utilized for backprojection, is a more suitable approximation for describing the visible volume of the RF ToF sensor.

However, to maintain consistency with the parameters given for camera-based systems, we assume an approximated 53° field of view in Table 1 of the main paper, which encompasses all of the extents measured with the FWHM and is highlighted in blue in Figure 10.

14 Dataset Post-processing and Evaluation

Example images of all 45 objects in MAROON can be found in Figure 20. We compare the reconstructions produced by the four presented depth imagers with a ground-truth reconstruction in a common metric space and describe the methods used in this process.

Projection into 3D. We acquire a point cloud of the object's surface utilizing the 2D depth and auxiliary data provided by MAROON. For a given pixel position (u, v) and its corresponding depth d from an optical depth sensor, we first verify its validity using the segmentation map of the same resolution — a step that has already been performed for radar during depth filtering. Subsequently we project each valid triple (u, v, d) back into 3D space using the given transformation matrix $T \in \mathbb{R}^{4 \times 4}$:

$$\begin{pmatrix} x \\ y \\ d \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} & I & & t \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1}}_{T^{-1}} \begin{pmatrix} u \cdot a \\ v \cdot a \\ d \\ 1 \end{pmatrix}. \tag{19}$$

For all optical depth imagers, this equation is the inverse of a perspective transformation with intrinsic camera matrix $I \in \mathbb{R}^{3 \times 3}$, pixel offset vector $t \in \mathbb{R}^3 = 0$ and $a = d$. Analogously for radar data, the equation is the inverse of an orthographic transformation with a scale matrix I, pixel offset t, and $a = 1$.

Joint Alignment. To estimate the deviation of a sensor reconstruction $R_s \in \mathbb{R}^{M\times3}$ from the GT, $R_g \in \mathbb{R}^{N\times3}$, we use the previously determined spatial calibration parameters $K_{g\to s} \in \mathbb{R}^{4\times4}$ to transform R_g from the GT space g into the sensor space s:

$$\widetilde{R}_g^s = \widetilde{R}_g K_{g\to s}^T . \tag{20}$$

$\widetilde{R}$ denotes the homogeneous version of R. We use the notation R^* to indicate a reconstruction that has been transformed to sensor space $*$.

14.1 Radar Depth Filtering

For our experiments and dataset post-processing, we chose an empirical threshold of -14 dB over all objects, which — to the best of our knowledge — has proven to yield the best balance of noise pruning while retaining relevant object measurements. In Figure 11, we show how the signal-to-noise ratio of the radar confidence map, i.e., the pixel-wise values of κ, behaves over different thresholds for the exemplary capture of the *S1 Hand Open.*
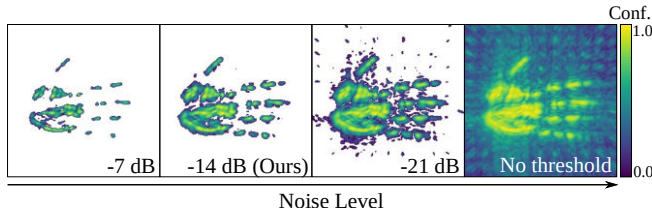

Fig. 11. Visualization of the 2D confidence map of the *S1 Hand Open* at various thresholds. The filtered confidence map is subsequently used to extract valid depth information.

Threshold	C_g $(w=2)$	C_s $(w=1)$	P $(w=1)$	P_e $(w=1)$	Weighted Mean
- 7 dB	1.37	**0.67**	**0.78**	**0.76**	0.99
-10 dB	1.07	*0.73*	*0.82*	*0.81*	0.90
-14 dB	0.82	0.9	0.95	0.85	**0.87**
-17 dB	0.68	1.05	1.03	0.95	*0.88*
-21 dB	*0.51*	1.54	1.16	1.07	0.96
—	**0.35**	5.75	1.45	1.45	1.87

Table 6. Ablation study with different signal thresholds used for depth filtering. The depth deviation is expressed in centimeters across the four metrics presented in the main paper, averaged for all objects at a 30 cm object-to-sensor distance. Additionally, we provide a weighted mean for each row, assigning double the weight, w, to C_g, as it is most sensitive to point cloud completeness. The **best** and *second best* results per metric are highlighted.

Additionally, we performed an ablation study to evaluate how different thresholds impact the mean depth deviation across all MAROON objects at a 30 cm object-to-sensor distance. In Table 6, we present results for the four metrics discussed in the main paper. We include a weighted mean for each row, giving double weight to C_g, as it is the only metric that is sensitive to the completeness of

the point cloud. Notably, we find that the performance concerning C_g is inversely related to that of the other metrics, which are more sensitive to signal noise and depth quality. The best trade-off between completeness and noise is achieved with a threshold ranging from -14 dB to -17 dB.

14.2 Radar Material Classification

To investigate the radar signal response and depth deviation with respect to different materials, we divided the 45 objects of MAROON into six classes. These assignments are listed in Table 9. The goal of this classification is to highlight material differences on a coarse level, noting the large object variety that still persists within one material class. Furthermore, we list the objects that are larger than the antenna aperture. It is important to consider these objects when interpreting the depth deviation trends presented in the main paper, as their reconstructions may be incomplete due to the portions that fall outside the antenna aperture.

14.3 Additional Results

We provide additional quantitative results for all 45 objects with respect to the depth deviation from the ground truth in Table 10, Table 11, Table 12, and Table 13.

15 Extended Discussion of Depth Deviation

First, we present complementary perspectives on the data, where we put the depth deviation of all 45 objects into relation with the different metric types and, subsequently, the different types of depth imagers. We analyze each representation in turn, highlighting common trends, in combination with previously stated results of Section 6.2.

15.1 General Trends

Interpreting the extensive numerical data on depth deviations in Table 10, Table 11, Table 12, and Table 13 can be challenging, so we provide visual, complementary views in this section. In order to relate different quantities to each other, we use barycentric interpolation based on triples of metric types (Figure 12) and sensors (Figure 13), respectively. For each triple (μ_a, μ_b, μ_c), the mean values for depth deviation (μ) of each object are transformed to affine coordinates (w_a, w_b, w_c) by using the formula $w_{\{a,b,c\}} = \mu_{\{a,b,c\}}/(\mu_a + \mu_b + \mu_c)$. Circle locations closer to a triangle corner indicate higher relative depth deviation. Moreover, as the triples are drawn from a set of four, the triangles are arranged in the shape of an unfolded tetrahedron, highlighting that each triangle's contents can be seen as a projection of barycentric coordinates within a (3D) tetrahedron $w_{\{a,b,c,d\}} = w_{\{a,b,c,d\}}/(w_a + \cdots + w_d)$.

Interpretation of Metrics. In Figure 12, we provide a qualitative comparison of each sensor's depth deviation from GT with respect to the four presented metrics. In dense reconstructions, as is typical for optical depth sensors, metrics based on nearest neighbors (here, C_g and C_s) are bound to be lower than those based on projection (P and P_e); they also tend to be more resilient against noise. For RF reconstructions, however, that are prone to sparse depth maps, Chamfer distances often create false matches; accordingly,
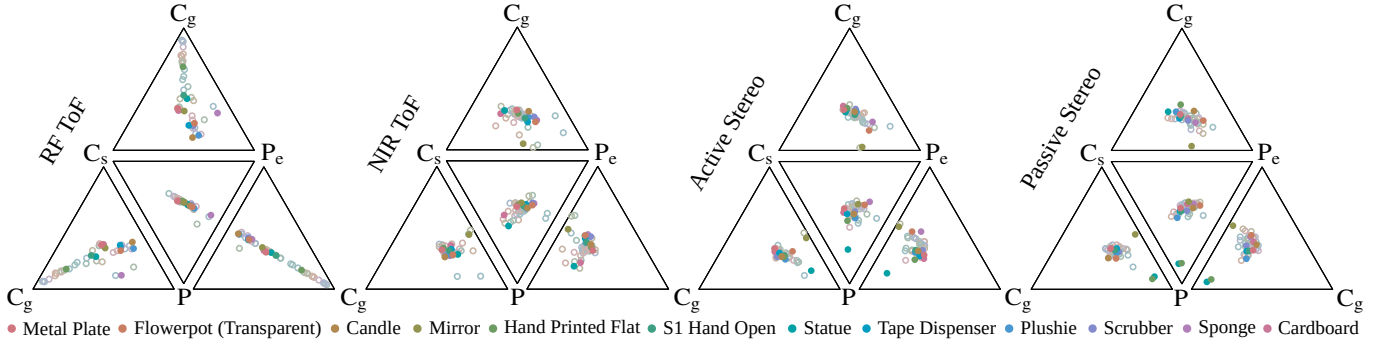
● Metal Plate ● Flowerpot (Transparent) ● Candle ● Mirror ● Hand Printed Flat ● S1 Hand Open ● Statue ● Tape Dispenser ● Plushie ● Scrubber ● Sponge ● Cardboard

Fig. 12. A complementary view on the depth deviation across different *metric types*. For each triplet of metrics $M_i \in \{C_g, C_s, P, P_e\}$, we convert each mean depth deviation μ_i to affine coordinates, $\bar{\mu}_i = \mu_i / \sum_i \mu_i$, that map an object's errors in to a triangle whose corners correspond to metrics M_i. All 45 MAROON objects are shown as circles, with selected objects from Table 4 highlighted in solid colors. Samples closer to a triangle corner indicate a higher relative depth deviation in the corresponding metric.

● Metal Plate ● Flowerpot (Transparent) ● Candle ● Mirror ● Hand Printed Flat ● S1 Hand Open ● Statue ● Tape Dispenser ● Plushie ● Scrubber ● Sponge ● Cardboard
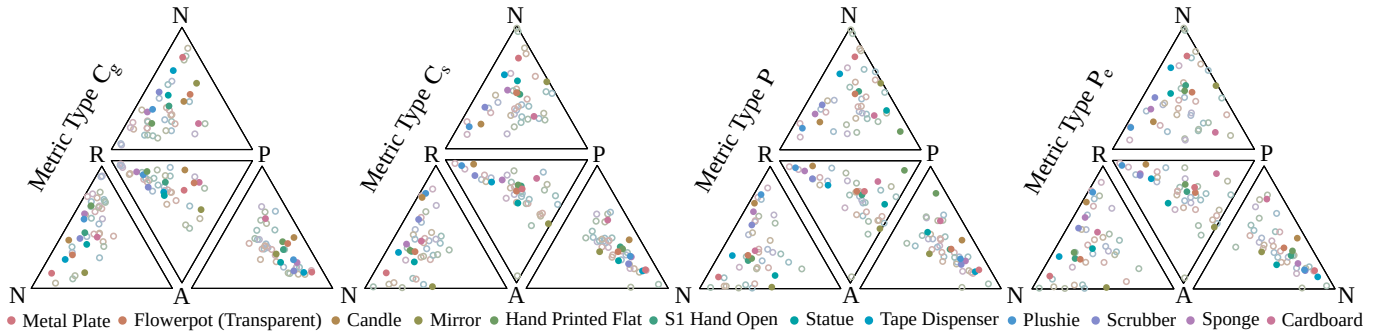
Fig. 13. A complementary view on the depth deviation across different *sensors*. The sensors are denoted as **R** (RF ToF), **N** (NIR ToF), **A** (Active Stereo), and **P** (Passive Stereo), respectively. Analogously to Figure 12, we convert the mean depth deviations μ_i to affine coordinates within triangles corresponding to all possible sensor triples. All 45 MAROON objects are shown as circles, with selected objects from Table 4 highlighted in solid colors. Samples closer to a triangle corner indicate a higher relative depth deviation for the corresponding sensor.
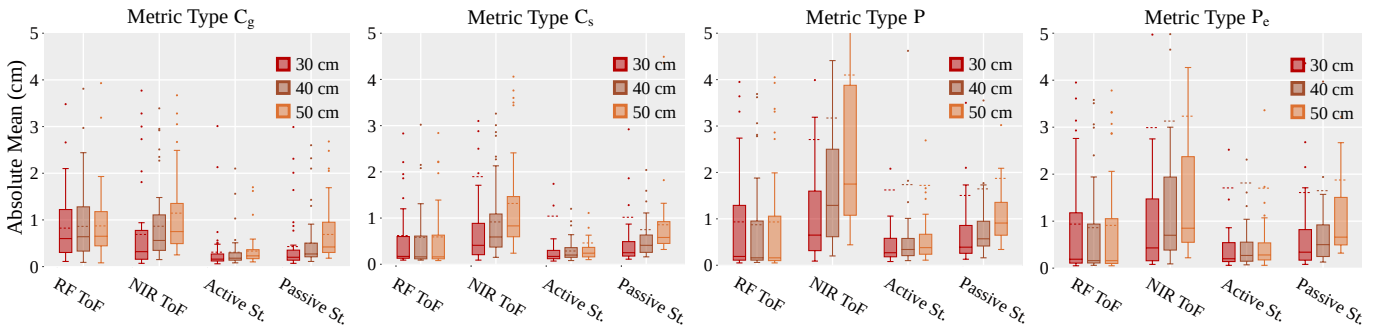

Fig. 14. Box plots, visualizing the distribution of the mean error across all objects with respect to different object-to-sensor distances. Solid (—) and dashed (- -) horizontal lines indicate the median and the mean of the distribution, respectively. The results are discussed in Section 15.

C_g dominates for RF ToF compared to the other metrics. For further discussion regarding the sparsity of RF reconstructions, see Section 7.3 in the main paper.

For optical sensors, a marginal trend towards the corners of C_s and P, away from the silhouette-resilient P_e and C_g, cf. Table 3, indicates the presence of noise at object silhouettes.

Relative Depth Deviation across Sensors. In Figure 13, we observe a considerable spread of depth deviations across different sensors. As noted in Section 6.2, the relative depth deviation between sensors ranges from 1.9 to 3.4 mm, that is, the variation in the depicted normalized error occurs within a comparatively small range of absolute errors. As a general trend, the two stereo sensors have the

lowest depth deviation (see triangles N–R–A and N–R–P), with a moderate edge for active stereo, particularly for metrics P and P_e, see the A–P axis, where passive stereo's performance degrades. This is consistent with the results of Section 6.2.1 in the main paper, where we find that the active stereo has the highest number of best results. However, given that relative depth deviations between sensors differ by only a few millimeters, we conclude that the seemingly improved depth quality of active stereo sensors is of minimal significance. Moreover, an examination of the scatter plot reveals a marginal weight trend towards the corners of the two ToF sensors. Reconstruction errors between the two ToF sensors are highly object-dependent, and many objects, including many of the highlighted ones, demonstrate multi-path effects due to perfect signal reflections, retro-reflectivity, or partial signal transmission. Further details on this will be discussed in the next section.

Depth Deviation over Distance. In Figure 14, we observe that the depth deviation of the passive stereo and the NIR ToF sensor is considerably more distance-specific compared to active stereo and RF ToF. For passive stereo, this may be due to a decrease in effective spatial resolution, where δ_z directly depends on $\delta_{x,y}$, see Section 11. As the distance between the object and the sensor increases, $\delta_{x,y}$ decreases, resulting in a loss of high-frequency color details while the object appears smaller in the image. Compared to passive stereo, the active stereo sensor has a comparably higher effective resolution, assuming that the resolution of both sensors differs in accordance with δ_{xyz} in Table 1 of the main paper. The unique active NIR pattern may also be less sensitive to decreases in spatial resolution, maintaining the quality of correspondence matches. The trend for the NIR ToF sensor aligns with findings from Bamji et al. [2018] for 30–50 cm distances. However, we argue that absolute errors for a target with 20% reflectivity do not fully represent all objects in our experiments. We suggest that, in addition to the expected decrease in spatial resolution, a greater depth deviation with increasing distance arises from the signal-to-noise ratio with respect to environmental light, which typically decreases over distance due to the inverse-square law.

Notably, RF ToF does not exhibit a distance-dependent depth deviation, unlike its optical counterpart. This seems to contradict Table 1 of the main paper, where δ_z for RF ToF degrades more rapidly with depth, compared to optical sensors; however, the theoretical decay with depth stems from the worsening separability of neighboring point targets, which is not a pertinent scenario in our database, as the recorded targets primarily have smooth surfaces at locations where valid reconstructions are measured. Moreover, our setup minimizes mmWave interactions with external objects, limiting noise primarily to the object itself. As a result, the signal-to-noise ratio of RF ToF is considerably less sensitive to changes in object-to-sensor distance compared to NIR ToF, assuming no interference from external sources.

16 Discussion of the Influence of Geometry on Reconstruction Completeness and Depth Deviation

In Section 7.3 of the main paper, we discussed the influence of object geometry on the RF ToF sensor. Here, we extend our experiments to all four depth imagers.

Influence on Reconstruction Completeness. We visualize the extended results for all depth imagers in Figure 15, where we show the mean depth deviation with respect to C_g in conjunction with the median surface incidence angle. We generally observe lower errors for optical sensors compared to RF ToF, indicating that optical depth measurements tend to be more complete.

During dataset capture, we aligned the object surfaces with the RF ToF sensor aperture. Due to spatial constraints, all optical sensors were placed next to the antenna aperture, resulting in view directions that do not directly align with the majority of surface normals. This is further illustrated in Figure 16 on the *right*, which depicts the sensor placement. Consequently, the object measurements from optical sensors begin at a median surface incidence angle of approximately 20-30°. The absence of a significant proportion of smaller angles in the data complicates the identification of notable trends.

Influence on Depth Deviation. We conduct an additional experiment measuring per-angle depth deviation with respect to metric P, which is generally more sensitive to depth quality and noise than C_g. For each object, we compute the surface incidence angle and depth deviation for each point-wise measurement using the corresponding ground-truth normal. We then aggregate the point-wise measurements across all objects and cluster them into angle bins of 5°. After calculating the mean depth deviation for each angle bin, we normalize the results to $[0, 1]$ across all four sensors.

The findings are depicted in Figure 16, where the length of each bar represents the relative quantity of per-point measurements for each angle bin, offering insight into data distribution. We illustrate the sensor setup on the *right* to clarify each sensor's placement and viewing direction, from which we derive the surface incidence angle.

As previously noted for the median angle measurements, the main lobe of per-angle measurements is concentrated around 30° for optical sensors and 0° for the RF ToF sensor, due to the respective sensor placements. In general, we observe a more rapid decline in depth quality for active sensors (NIR ToF, active stereo, RF ToF) compared to passive stereo, which underscores their dependency on well-illuminated (or well-radiated) areas. Additionally, the depth quality of the RF ToF sensor significantly decreases at angles greater than 30°, rendering it more susceptible to object geometry than optical sensors, where we experience a decline in depth quality at angles greater than 60°.

17 Extended Discussion of MIMO Radar: Signal Response and Depth Deviation

In this section, we extend the experiments of Section 7.3 of the main paper, where we presented both, the signal magnitude and the depth deviation concerning object material, geometry, and size. To demonstrate that the signal magnitude is not only influenced by either object geometry or by object size — which would be possible due to the high diversity of captured objects that prevents us from isolating one variable while keeping the others constant — we visualize the signal magnitude in Figure 17 (*top*), in relation to object geometry (median surface incidence angle) and size (relative surface area).

A general trend on the x-axis shows an increase in the signal magnitude from left to right, particularly for objects with a surface
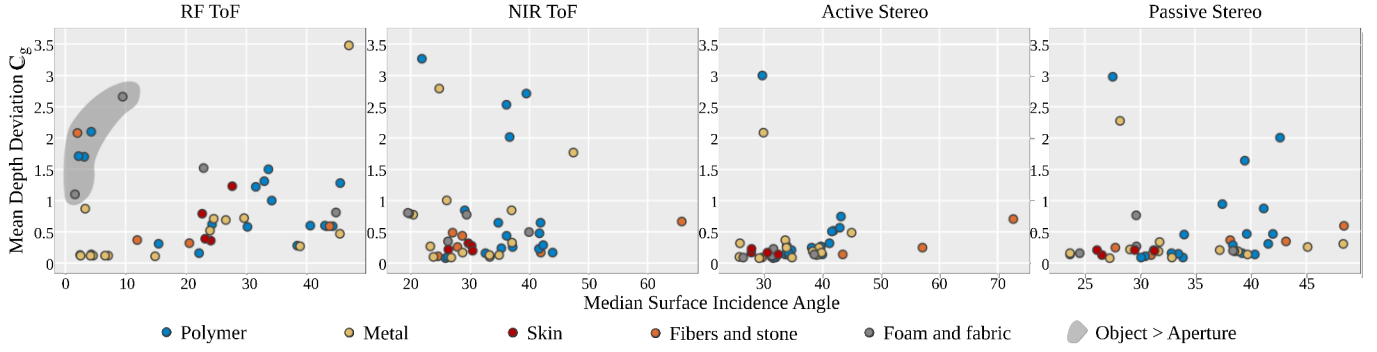
Fig. 15. Extended experiments for all four depth imagers, where object material, geometry (median surface incidence angle), and size (relative surface area) is put in relation to mean depth deviation. Measurements, where large objects appear outside the radar's antenna aperture, are highlighted in gray regions, as they exhibit higher depth deviations compared to the ground-truth reconstructions, which may extend beyond this aperture; this is attributed to the comparably small field of view and the surface reflection characteristics with respect to radio waves (see Section 13.1).
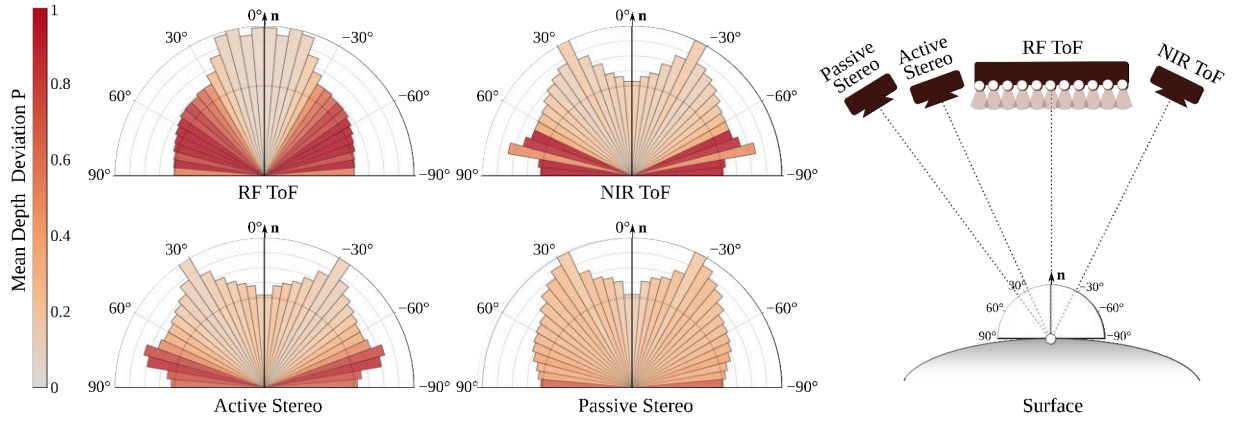

Fig. 16. Depth deviation with respect to metric P per 5°-binned surface incidence angle, normalized to $[0, 1]$ across the four presented depth imagers. The length of each bin indicates the relative data distribution, with the minimum bin length represented as a dotted hemisphere contour. The corresponding sensor setup is shown on the *right*, providing an intuitive understanding of the data distribution in relation to each sensor's viewing direction and the majority of ground-truth surface normals pointing into the direction of $\boldsymbol{n}$.

incidence angle greater than $10°$. The majority of objects below this $10°$ angle on the y-axis exhibit significantly higher signal magnitudes, regardless of their relative surface area, hence indicating the influence of object geometry. On the *bottom* of Figure 17, we visualize the mean depth deviation in relation to object geometry and size. While a similar trend is observed for object geometry on the y-axis — where objects of lower surface incidence angle exhibit smaller errors — no overall trend appears on the x-axis, suggesting depth deviation is more influenced by object geometry instead of size.

17.1 Ablation Study with Reduced Antenna Architectures

To explore alternative consumer-friendly RF ToF devices, we experiment with different architectures, varying the aperture size and antenna density by selecting only a subset of antennas from the raw

phasor measurements in MAROON. Two of the selected antenna architectures are shown in Figure 18, both with a comparable number of antennas.

We visually compare the reconstructions of four antenna configurations in Figure 19 for the *Hand Printed Flat* at 30 cm object-to-sensor distance. Reducing the aperture size (*right column*) results in a loss of continuous object geometry, causing the object to visually resemble a collection of point targets. Compared to architectures with larger apertures (*left column*), less of the surface details are preserved. Conversely, a decreasing antenna density introduces more localization ambiguities, leading to multiple reconstructions of the object, as seen with the replicated hands in the *lower left* reconstruction.

Further quantitative results are presented in Table 7, where we measure the mean depth deviation in centimeters across all objects at a distance of 30 cm while varying the antenna architecture. When comparing configurations with similar numbers of antennas, larger
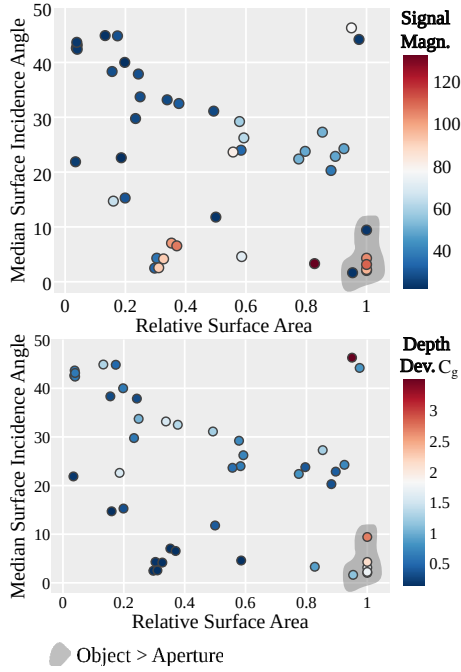
Fig. 17. The mean signal magnitude (*top*) and mean depth deviation (*bottom*), put in relation to object geometry (median surface incidence angle) and object size (relative surface area). Large objects outside the radar's antenna aperture exhibit higher depth deviations with respect to the ground-truth reconstructions that possibly extend beyond this aperture; this is attributed to the comparably small field of view and the surface reflection characteristics with respect to radio waves (see Section 13.1).
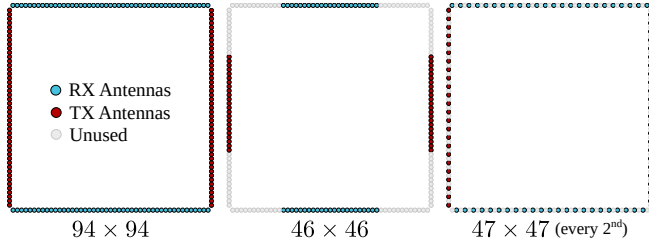

Fig. 18. We simulate more consumer-friendly antenna apertures and compare them to the full 94×94 antenna array. First, we reduce the aperture size by selecting a spatially centered antenna subset within the MIMO array. Second, we reduce the antenna density, using only every second antenna from the array, while preserving an aperture size comparable to that of the original array. Additionally, we maintain a similar number of antennas between both consumer-friendly variants to ensure a fair comparison.

aperture sizes exhibit lower depth deviation than higher-density configurations, as indicated by P and P_e; this supports prior qualitative observations that surface quality declines more rapidly with reduced aperture size. In contrast, decreasing the antenna density leads to a rapid rise in noise, likely due to the previously mentioned
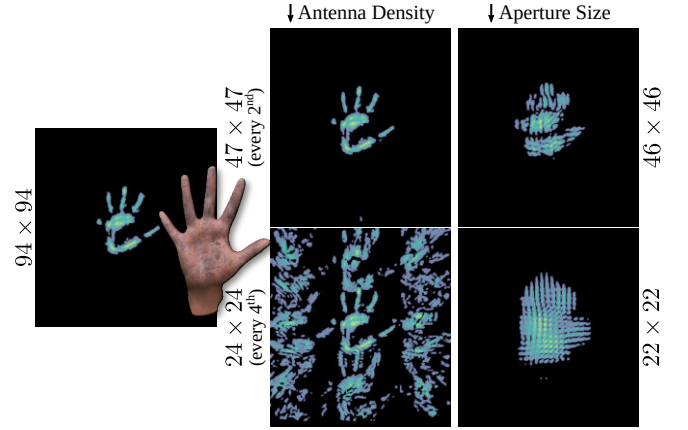

Fig. 19. Qualitative evaluation of different antenna configurations for the *Hand Printed Flat* object at 30 cm object-to-sensor distance, using only 50 % and 25 % of the original number of antennas.

localization ambiguities, which tend to appear in areas where holes typically arise, as C_g behaves inversely proportional to C_s.

Antenna Config.	C_g	C_s	P	P_e
46×46	1.01	**1.09**	*1.09*	*1.08*
22×22	0.82	1.59	1.30	1.24
47×47 (every 2nd)	*0.73*	*1.2*	**1.02**	**0.95**
24×24 (every 4th)	**0.59**	5.97	1.41	1.35
94×94	0.82	0.90	0.94	0.85

Table 7. Ablation study on different antenna configurations. The depth deviation is expressed in centimeters across the four metrics presented in the main paper, averaged for all objects at a 30 cm object-to-sensor distance. The **best** and *second best* results per metric are highlighted.

17.2 Ablation Study with Reduced Frequencies

In this experiment, we adjust the frequency configuration of the RF ToF sensor to simulate various sensors. For this, we compute frequency subsets of the raw phasor data provided in the dataset. Following a similar approach to the antenna aperture ablations described in Section 17.1, we implement two key variations: first, by operating with a smaller bandwidth, and second, by varying the frequency differences through subsampling every second, fourth, eighth frequency, and so on.

The results are shown in Table 8, with mean depth deviations quantified in centimeters across all metrics and objects at a distance of 30 cm while varying the frequency configuration. Compared to the full frequency spectrum used in the main paper (*last row*), the 64-frequency-stepped configuration performs on par, indicating that a configuration with half the frequencies may serve as a viable alternative maintaining the same accuracy.

Additionally, we observe a general trend of increasing depth deviation with a decreasing number of frequencies. Among the variations,

adjusting the frequency difference rather than the bandwidth yields better overall results and may present an interesting sensor configuration that could enhance the RF ToF sensor's capture rate — as fewer frequencies require less capture time.

Frequency Configuration (in GHz)	C_g	C_s	P	P_e
$f_m \in [72.00, 81.92]$, $\Delta f = 0.16$, $N_f = 64$	*0.81*	*0.91*	*0.95*	*0.86*
$f_m \in [72.00, 81.76]$, $\Delta f = 0.31$, $N_f = 32$	0.90	1.56	1.24	1.15
$f_m \in [72.00, 81.45]$, $\Delta f = 0.59$, $N_f = 16$	0.90	1.56	1.24	1.16
$f_m \in [72.00, 80.82]$, $\Delta f = 1.10$, $N_f = 8$	0.86	3.58	2.49	2.20
$f_m \in [72.00, 79.56]$, $\Delta f = 1.89$, $N_f = 4$	0.91	4.82	4.2	4.04
$f_m \in [76.96, 82.00]$, $\Delta f = 0.078$, $N_f = 64$	**0.76**	1.19	1.21	1.11
$f_m \in [79.48, 82.00]$, $\Delta f = 0.078$, $N_f = 32$	**0.76**	1.98	1.85	1.67
$f_m \in [80.74, 82.00]$, $\Delta f = 0.078$, $N_f = 16$	0.77	3.33	2.87	2.71
$f_m \in [81.37, 82.00]$, $\Delta f = 0.078$, $N_f = 8$	0.91	5.15	4.14	3.97
$f_m \in [81.69, 82.00]$, $\Delta f = 0.078$, $N_f = 4$	0.98	6.22	4.82	4.71
$f_m \in [72.00, 82.00]$, $\Delta f = 0.078$, $N_f = 128$	0.82	**0.90**	**0.94**	**0.85**

Table 8. Ablation study on various frequency configurations defined by the range of the modulation frequency f_m, frequency difference Δf, and the corresponding number of frequency steps N_f. Frequency units are given in GHz. The *last row* depicts the full-bandwidth frequency configuration from the main paper. The depth deviation is expressed in centimeters, averaged for all objects at a 30 cm object-to-sensor distance. The **best** and *second best* results per metric are highlighted.

Cardboard	Metal Disk (Thin)	Metal Disk (Thick)	Concrete Stone	Sponge	Wood Ball
Scrubber	Cardboard Box	Plushie	Bottle	Tape Dispenser	Book
Statue	Rubber Foam Plane	Sandpaper (k80)	Sandpaper (k120)	Wood Plane	Foam Plane
S1 Hand Open	S1 Hand Open (Rev.)	S2 Hand Open	S2 Hand Open (Rev.)	Hand Printed Flat	Corner Reflector
Mirror	Candle	Flowerpot (Transparent)	V1 Metal Plate	V2 Metal Plate	Hand Printed F
Hand Printed B	Hand Printed U	Metal Angle	Plunger	Silicone Cup	V1 Christmas Ball
V2 Christmas Ball	V3 Christmas Ball	Water Cube	Flowerpot (Brown)	Brazen Rosette	Pool Ball
Polystyrene Plate	Bunny	Bunny Box			

Fig. 20. Example images of all objects in MAROON.

Class	Objects	Additional Description
Metal	V1 Metal Plate, V2 Metal Plate Metal Disk (Thin), Metal Disk (Thick) Hand Printed: Flat, B, F, U Brazen Rosette Corner Reflector Cardboard Box Mirror Metal Angle Statue	Coated with metal lacquer Coated with metal lacquer Metal surface beneath partially transmissive glass
Fibers and stone	Wood Plane Cardboard Book Concrete Stone Wood Ball Bunny Box	Primarily made out of paper Large wooden box, comparably small plastic bunny
Polymer	Plumber Silicone Cup Christmas Ball: V1, V2, V3 Candle Bottle Sandpaper (k120), Sandpaper (k80) Flowerpot (Transparent), Flowerpot (Brown) Polystyrene Plate Water Cube Scrubber Pool Ball Bunny Tape Dispenser	Water wrapped in a plastic cube
Skin	S1 Hand Open, S1 Hand Open (Rev.) S2 Hand Open, S2 Hand Open (Rev.)	
Foam and fabric (Primarily transmissive)	Rubber Foam Plane Sponge Plushie Foam Plane	
Objects outside FOV	Polystyrene Plate Sandpaper (k120), Sandpaper (k80) Wood Plane Foam Plane Rubber Foam Plane	

Table 9. Assignment from objects to material classes with optional description about the assignment process.

	Metric Type	Cardboard	Metal Disk (Thin)	Metal Disk (Thick)	Concrete Stone	Sponge	Wood Ball
RF ToF	C_g	0.13 (± 0.06)	<u>0.12</u> (± <u>0.04</u>)	<u>0.12</u> (± **0.05**)	<u>0.14</u> (± <u>0.07</u>)	<u>1.52</u> (± <u>0.97</u>)	<u>0.59</u> (± <u>0.40</u>)
NIR ToF		0.10 (± 0.06)	0.09 (± **0.03**)	0.08 (± **0.05**)	0.09 (± **0.04**)	0.79 (± 0.45)	0.17 (± 0.12)
Active Stereo		**0.08** (± 0.05)	**0.06** (± <u>0.04</u>)	**0.07** (± 0.05)	**0.08** (± 0.04)	**0.18** (± 0.17)	**0.13** (± **0.06**)
Passive Stereo		<u>0.24</u> (± <u>0.11</u>)	0.07 (± <u>0.04</u>)	0.08 (± <u>0.07</u>)	0.09 (± 0.06)	0.26 (± **0.15**)	0.34 (± 0.16)
RF ToF	C_s	0.15 (± 0.08)	<u>0.13</u> (± **0.05**)	0.13 (± **0.06**)	<u>0.15</u> (± 0.08)	0.54 (± 0.40)	**0.12** (± **0.05**)
NIR ToF		0.16 (± <u>0.18</u>)	0.12 (± 0.13)	<u>0.14</u> (± <u>0.20</u>)	0.10 (± 0.06)	<u>0.79</u> (± <u>0.42</u>)	0.21 (± 0.19)
Active Stereo		**0.08** (± 0.05)	**0.07** (± 0.05)	**0.08** (± 0.07)	**0.08** (± **0.04**)	**0.17** (± 0.15)	0.13 (± 0.07)
Passive Stereo		<u>0.30</u> (± 0.13)	0.12 (± <u>0.17</u>)	<u>0.14</u> (± <u>0.33</u>)	0.12 (± <u>0.08</u>)	0.33 (± 0.18)	<u>0.40</u> (± <u>0.20</u>)
RF ToF	P	0.13 (± 0.14)	0.10 (± **0.11**)	0.11 (± **0.14**)	0.13 (± 0.15)	<u>2.73</u> (± <u>2.47</u>)	**0.10** (± **0.09**)
NIR ToF		0.19 (± <u>0.25</u>)	0.13 (± 0.19)	0.16 (± 0.28)	**0.10** (± **0.10**)	1.32 (± 0.58)	0.32 (± <u>0.44</u>)
Active Stereo		**0.10** (± 0.12)	**0.08** (± 0.25)	**0.10** (± 0.17)	**0.10** (± 0.12)	**0.29** (± 0.42)	0.20 (± 0.17)
Passive Stereo		<u>0.36</u> (± 0.17)	<u>0.14</u> (± <u>0.32</u>)	<u>0.18</u> (± <u>0.51</u>)	0.16 (± 0.17)	0.47 (± 0.51)	<u>0.65</u> (± <u>0.33</u>)
RF ToF	P_e	0.12 (± 0.13)	0.10 (± 0.08)	<u>0.10</u> (± 0.09)	<u>0.13</u> (± <u>0.15</u>)	2.93 (± 2.60)	**0.10** (± 0.09)
NIR ToF		0.08 (± 0.10)	<u>0.12</u> (± **0.06**)	0.09 (± **0.06**)	**0.09** (± **0.07**)	1.69 (± **0.23**)	**0.10** (± 0.10)
Active Stereo		**0.06** (± **0.08**)	**0.06** (± 0.06)	**0.06** (± 0.06)	0.09 (± 0.10)	**0.42** (± 0.50)	0.16 (± **0.07**)
Passive Stereo		<u>0.38</u> (± <u>0.14</u>)	0.08 (± <u>0.11</u>)	0.08 (± <u>0.10</u>)	0.10 (± 0.10)	0.55 (± 0.56)	<u>0.69</u> (± <u>0.26</u>)
RF ToF	P^*	+0.08 (± 0.14)	-0.06 (± **0.11**)	-0.05 (± **0.14**)	-0.06 (± 0.15)	<u>+2.25</u> (± <u>2.47</u>)	**-0.07** (± **0.09**)
NIR ToF		+0.15 (± <u>0.25</u>)	-0.04 (± 0.19)	**-0.01** (± 0.28)	-0.08 (± **0.10**)	+1.30 (± 0.58)	+0.09 (± <u>0.44</u>)
Active Stereo		**+0.04** (± 0.12)	**-0.01** (± 0.25)	-0.02 (± 0.17)	**-0.04** (± 0.12)	**-0.15** (± 0.42)	+0.17 (± 0.17)
Passive Stereo		<u>+0.36</u> (± 0.17)	<u>+0.07</u> (± <u>0.32</u>)	<u>+0.10</u> (± <u>0.51</u>)	<u>+0.15</u> (± <u>0.17</u>)	+0.22 (± 0.51)	<u>+0.64</u> (± <u>0.33</u>)
RF ToF	P_e^*	+0.08 (± 0.13)	-0.09 (± 0.08)	<u>-0.08</u> (± 0.09)	-0.07 (± <u>0.15</u>)	<u>+2.40</u> (± <u>2.60</u>)	-0.07 (± 0.09)
NIR ToF		**-0.00** (± 0.10)	<u>-0.12</u> (± **0.06**)	-0.08 (± **0.06**)	<u>-0.08</u> (± **0.07**)	+1.69 (± **0.23**)	-0.05 (± 0.10)
Active Stereo		**+0.00** (± **0.08**)	-0.04 (± **0.06**)	-0.05 (± **0.06**)	**-0.05** (± 0.10)	-0.37 (± 0.50)	+0.16 (± **0.07**)
Passive Stereo		<u>+0.38</u> (± <u>0.14</u>)	**+0.01** (± <u>0.11</u>)	**+0.02** (± <u>0.10</u>)	<u>+0.08</u> (± 0.10)	**-0.34** (± 0.56)	<u>+0.69</u> (± <u>0.26</u>)

	Metric Type	Scrubber	Cardboard Box	Plushie	Bottle	Tape Dispenser	Book
RF ToF	C_g	0.58 (± 0.29)	0.12 (± 0.08)	<u>0.81</u> (± <u>0.47</u>)	<u>1.22</u> (± <u>1.00</u>)	0.31 (± 0.23)	0.37 (± 0.56)
NIR ToF		<u>0.64</u> (± <u>0.34</u>)	**0.11** (± **0.04**)	0.49 (± 0.21)	0.44 (± 0.51)	<u>0.84</u> (± <u>0.30</u>)	<u>0.59</u> (± <u>0.65</u>)
Active Stereo		0.20 (± 0.16)	0.12 (± 0.07)	**0.13** (± **0.14**)	**0.16** (± **0.21**)	0.16 (± 0.14)	**0.13** (± **0.11**)
Passive Stereo		**0.14** (± **0.11**)	<u>0.29</u> (± <u>0.20</u>)	0.19 (± 0.21)	0.28 (± 0.36)	**0.15** (± **0.11**)	0.15 (± 0.13)
RF ToF	C_s	<u>0.97</u> (± <u>0.61</u>)	**0.11** (± **0.05**)	<u>2.21</u> (± <u>2.05</u>)	0.49 (± 0.99)	0.55 (± <u>0.77</u>)	**0.14** (± **0.08**)
NIR ToF		0.59 (± 0.30)	0.14 (± 0.16)	0.53 (± 0.28)	<u>0.80</u> (± <u>1.32</u>)	<u>1.16</u> (± 0.60)	0.20 (± 0.17)
Active Stereo		**0.17** (± **0.11**)	0.15 (± 0.13)	**0.12** (± 0.38)	**0.20** (± **0.53**)	**0.17** (± **0.17**)	0.18 (± 0.15)
Passive Stereo		0.19 (± 0.13)	<u>0.35</u> (± <u>0.24</u>)	0.23 (± **0.27**)	0.41 (± 0.62)	0.22 (± 0.20)	<u>0.24</u> (± <u>0.19</u>)
RF ToF	P	<u>1.28</u> (± <u>0.76</u>)	0.38 (± <u>1.52</u>)	<u>3.64</u> (± <u>3.23</u>)	0.64 (± 1.63)	0.67 (± <u>1.08</u>)	**0.13** (± 0.31)
NIR ToF		0.91 (± 0.47)	**0.14** (± **0.19**)	0.85 (± **0.52**)	<u>1.32</u> (± 2.00)	<u>1.57</u> (± 0.70)	<u>1.15</u> (± <u>1.42</u>)
Active Stereo		**0.27** (± **0.34**)	0.19 (± 0.22)	**0.24** (± 1.25)	**0.49** (± <u>2.49</u>)	**0.24** (± **0.35**)	0.19 (± **0.25**)
Passive Stereo		0.29 (± 0.38)	<u>0.47</u> (± 0.55)	0.35 (± 0.54)	0.62 (± **0.97**)	0.29 (± 0.35)	0.47 (± 1.21)
RF ToF	P_e	<u>1.35</u> (± 0.66)	0.25 (± 1.19)	<u>3.61</u> (± <u>3.23</u>)	0.63 (± 1.63)	0.66 (± <u>1.07</u>)	**0.13** (± 0.30)
NIR ToF		1.16 (± **0.27**)	**0.13** (± 0.07)	0.82 (± 0.38)	<u>1.16</u> (± 2.16)	<u>1.70</u> (± 0.89)	<u>1.34</u> (± <u>1.54</u>)
Active Stereo		**0.22** (± **0.27**)	0.18 (± 0.12)	**0.16** (± 0.30)	**0.33** (± 1.13)	**0.15** (± **0.21**)	0.17 (± **0.23**)
Passive Stereo		0.24 (± 0.29)	<u>0.42</u> (± 0.28)	0.18 (± **0.26**)	0.55 (± **0.86**)	0.20 (± 0.22)	0.33 (± 0.95)
RF ToF	P^*	<u>+1.26</u> (± <u>0.76</u>)	+0.28 (± <u>1.52</u>)	<u>+3.63</u> (± <u>3.23</u>)	+0.52 (± 1.63)	+0.56 (± <u>1.08</u>)	-0.10 (± 0.31)
NIR ToF		+0.89 (± 0.47)	**-0.09** (± **0.19**)	+0.80 (± **0.52**)	<u>+1.30</u> (± 2.00)	<u>+1.57</u> (± 0.70)	<u>+1.13</u> (± <u>1.42</u>)
Active Stereo		**-0.01** (± 0.34)	-0.10 (± 0.22)	**-0.04** (± 1.25)	**+0.18** (± <u>2.49</u>)	**+0.01** (± 0.35)	**-0.07** (± 0.25)
Passive Stereo		**-0.01** (± 0.38)	<u>+0.42</u> (± 0.55)	+0.17 (± 0.54)	+0.51 (± **0.97**)	+0.19 (± **0.35**)	+0.39 (± 1.21)
RF ToF	P_e^*	<u>+1.34</u> (± 0.66)	+0.14 (± <u>1.19</u>)	<u>+3.59</u> (± <u>3.23</u>)	+0.52 (± 1.63)	+0.57 (± <u>1.07</u>)	-0.10 (± 0.30)
NIR ToF		+1.16 (± **0.27**)	**-0.13** (± **0.07**)	+0.82 (± 0.38)	<u>+1.15</u> (± 2.16)	<u>+1.70</u> (± 0.89)	<u>+1.33</u> (± <u>1.54</u>)
Active Stereo		**-0.02** (± **0.27**)	-0.16 (± 0.12)	-0.06 (± 0.30)	**+0.21** (± 1.13)	**+0.03** (± **0.21**)	**-0.02** (± **0.23**)
Passive Stereo		-0.07 (± 0.29)	<u>+0.42</u> (± 0.28)	**+0.02** (± **0.26**)	+0.52 (± **0.86**)	+0.08 (± 0.22)	+0.27 (± 0.95)

Table 10. We measure the depth deviation with respect to C_g, C_s, P, P_e and an additional signed version of P,P_e, which is denoted as P^*,P_e^*. All metrics are listed in the form $(\mu \pm \sigma)$, consisting of the mean μ and standard deviation σ in centimeters, computed over the entire metric domain, respectively. The best results among all sensors of one metric type are highlighted in **bold** and the worst results are <u>underlined</u>.

	Metric Type	Statue	Rubber Foam Plane	Sandpaper (k80)	Sandpaper (k120)	Wood Plane	Foam Plane
RF ToF	C_g	0.27 (± 0.25)	<u>1.10</u> (± <u>1.21</u>)	<u>1.70</u> (± <u>2.07</u>)	<u>1.71</u> (± <u>2.23</u>)	<u>2.08</u> (± <u>2.31</u>)	<u>2.66</u> (± <u>1.26</u>)
NIR ToF		<u>0.32</u> (± <u>0.28</u>)	0.34 (± 0.07)	0.09 (± **0.04**)	**0.07** (± **0.03**)	0.48 (± 0.15)	0.80 (± 0.14)
Active Stereo		0.16 (± **0.13**)	**0.11** (± **0.05**)	**0.07** (± **0.04**)	0.08 (± 0.04)	0.13 (± **0.06**)	**0.08** (± **0.06**)
Passive Stereo		**0.13** (± **0.13**)	0.77 (± 0.67)	0.08 (± **0.04**)	0.10 (± 0.06)	**0.12** (± 0.10)	0.15 (± 0.09)
RF ToF	C_s	**0.17** (± **0.11**)	0.34 (± <u>0.72</u>)	<u>0.12</u> (± **0.05**)	<u>0.12</u> (± **0.05**)	0.20 (± 0.12)	<u>2.83</u> (± <u>0.89</u>)
NIR ToF		<u>0.43</u> (± 0.42)	0.38 (± 0.11)	0.11 (± <u>0.12</u>)	**0.09** (± <u>0.12</u>)	<u>0.52</u> (± <u>0.16</u>)	0.84 (± 0.15)
Active Stereo		0.19 (± <u>0.76</u>)	**0.13** (± **0.08**)	**0.08** (± **0.05**)	0.09 (± **0.05**)	0.15 (± **0.08**)	**0.09** (± 0.15)
Passive Stereo		0.18 (± 0.69)	<u>0.83</u> (± 0.71)	<u>0.12</u> (± 0.09)	0.11 (± 0.08)	**0.13** (± 0.11)	0.16 (± **0.12**)
RF ToF	P	**0.20** (± 0.26)	<u>1.62</u> (± <u>2.21</u>)	**0.09** (± 0.11)	**0.08** (± 0.10)	0.18 (± 0.23)	<u>3.95</u> (± <u>4.22</u>)
NIR ToF		0.77 (± 3.13)	0.44 (± **0.19**)	0.10 (± 0.14)	0.09 (± <u>0.14</u>)	<u>0.65</u> (± <u>0.28</u>)	1.13 (± 3.58)
Active Stereo		0.90 (± 5.17)	**0.16** (± 0.24)	**0.09** (± **0.09**)	0.10 (± **0.09**)	0.20 (± **0.12**)	0.21 (± 2.11)
Passive Stereo		<u>1.43</u> (± <u>7.09</u>)	0.91 (± 1.11)	<u>0.13</u> (± <u>0.17</u>)	<u>0.13</u> (± <u>0.14</u>)	**0.14** (± 0.18)	**0.19** (± **0.44**)
RF ToF	P_e	0.20 (± <u>0.27</u>)	<u>1.62</u> (± <u>2.21</u>)	0.09 (± 0.11)	**0.08** (± 0.10)	0.18 (± 0.23)	<u>3.95</u> (± <u>4.22</u>)
NIR ToF		<u>0.25</u> (± **0.13**)	0.43 (± 0.11)	**0.08** (± 0.10)	**0.08** (± 0.10)	<u>0.64</u> (± <u>0.27</u>)	0.91 (± 0.12)
Active Stereo		0.16 (± 0.21)	**0.16** (± **0.09**)	0.09 (± **0.08**)	0.10 (± **0.09**)	0.20 (± **0.10**)	**0.09** (± **0.10**)
Passive Stereo		**0.10** (± **0.13**)	0.94 (± 1.13)	<u>0.12</u> (± <u>0.15</u>)	<u>0.11</u> (± <u>0.12</u>)	**0.12** (± 0.13)	0.17 (± 0.20)
RF ToF	P^*	**-0.04** (± 0.26)	<u>-1.35</u> (± <u>2.21</u>)	-0.04 (± 0.11)	-0.04 (± 0.10)	**+0.03** (± 0.23)	-0.66 (± <u>4.22</u>)
NIR ToF		-0.23 (± 3.13)	+0.43 (± **0.19**)	-0.03 (± 0.14)	**+0.02** (± <u>0.14</u>)	<u>+0.65</u> (± <u>0.28</u>)	<u>+1.13</u> (± 3.58)
Active Stereo		+0.71 (± 5.17)	**-0.15** (± 0.24)	<u>-0.06</u> (± **0.09**)	<u>-0.08</u> (± **0.09**)	-0.19 (± **0.12**)	**+0.08** (± 2.11)
Passive Stereo		<u>+1.34</u> (± <u>7.09</u>)	+0.34 (± 1.11)	**+0.00** (± <u>0.17</u>)	<u>-0.08</u> (± <u>0.14</u>)	+0.08 (± 0.18)	**+0.08** (± **0.44**)
RF ToF	P_e^*	-0.04 (± <u>0.27</u>)	<u>-1.35</u> (± <u>2.21</u>)	-0.04 (± 0.11)	-0.04 (± 0.10)	**+0.03** (± 0.23)	-0.66 (± <u>4.22</u>)
NIR ToF		<u>-0.24</u> (± **0.13**)	+0.43 (± 0.11)	-0.04 (± 0.10)	**+0.00** (± 0.10)	<u>+0.64</u> (± <u>0.27</u>)	<u>+0.91</u> (± 0.12)
Active Stereo		+0.07 (± 0.21)	**-0.16** (± **0.09**)	<u>-0.06</u> (± **0.08**)	<u>-0.09</u> (± **0.09**)	-0.20 (± **0.10**)	**-0.05** (± **0.10**)
Passive Stereo		**+0.03** (± **0.13**)	+0.35 (± 1.13)	**+0.01** (± <u>0.15</u>)	-0.07 (± <u>0.12</u>)	+0.07 (± 0.13)	+0.07 (± 0.20)

	Metric Type	S1 Hand Open	S1 Hand Open (Rev.)	S2 Hand Open	S2 Hand Open (Rev.)	Hand Printed Flat	Corner Reflector
RF ToF	C_g	<u>0.36</u> (± <u>0.38</u>)	<u>1.23</u> (± <u>1.42</u>)	<u>0.39</u> (± <u>0.37</u>)	<u>0.79</u> (± <u>0.79</u>)	<u>0.71</u> (± <u>0.78</u>)	<u>3.48</u> (± <u>1.80</u>)
NIR ToF		0.31 (± 0.14)	0.27 (± 0.12)	0.19 (± **0.10**)	0.21 (± **0.11**)	0.25 (± 0.12)	1.81 (± 1.01)
Active Stereo		**0.12** (± **0.09**)	0.16 (± 0.15)	**0.16** (± 0.13)	0.21 (± 0.16)	**0.09** (± **0.07**)	0.48 (± 0.62)
Passive Stereo		0.20 (± 0.16)	**0.12** (± **0.09**)	0.20 (± 0.16)	**0.20** (± 0.15)	0.21 (± 0.37)	**0.30** (± **0.29**)
RF ToF	C_s	0.22 (± 0.15)	**0.17** (± **0.11**)	0.20 (± **0.14**)	**0.15** (± **0.09**)	0.17 (± 0.13)	1.95 (± **0.94**)
NIR ToF		<u>0.38</u> (± <u>0.26</u>)	<u>0.32</u> (± <u>0.24</u>)	<u>0.25</u> (± <u>0.23</u>)	<u>0.27</u> (± <u>0.24</u>)	<u>0.29</u> (± 0.20)	<u>1.97</u> (± 1.09)
Active Stereo		**0.13** (± **0.10**)	0.18 (± <u>0.73</u>)	0.17 (± 0.14)	0.22 (± 0.15)	**0.09** (± **0.06**)	0.55 (± <u>1.54</u>)
Passive Stereo		0.26 (± 0.22)	**0.17** (± 0.18)	<u>0.25</u> (± 0.20)	0.25 (± 0.19)	0.18 (± <u>0.34</u>)	0.61 (± 1.22)
RF ToF	P	**0.22** (± 0.25)	0.16 (± 0.21)	**0.20** (± 0.24)	0.14 (± 0.17)	**0.16** (± 0.20)	<u>3.31</u> (± **1.53**)
NIR ToF		<u>0.52</u> (± 0.43)	<u>0.47</u> (± 0.44)	<u>0.39</u> (± 0.57)	<u>0.35</u> (± 0.39)	0.33 (± 0.29)	2.87 (± 1.67)
Active Stereo		**0.22** (± <u>1.25</u>)	0.30 (± <u>1.54</u>)	0.29 (± <u>1.67</u>)	0.33 (± <u>1.09</u>)	**0.16** (± 1.30)	1.06 (± <u>3.92</u>)
Passive Stereo		0.35 (± 0.41)	0.22 (± 0.38)	0.35 (± 0.85)	0.32 (± 0.48)	<u>1.73</u> (± <u>8.95</u>)	**0.99** (± 2.49)
RF ToF	P_e	0.22 (± 0.25)	**0.16** (± 0.20)	**0.20** (± 0.24)	**0.14** (± **0.16**)	0.16 (± <u>0.20</u>)	<u>3.14</u> (± 1.41)
NIR ToF		<u>0.51</u> (± 0.27)	<u>0.38</u> (± 0.21)	<u>0.30</u> (± 0.22)	0.25 (± 0.17)	<u>0.30</u> (± **0.09**)	2.65 (± 1.29)
Active Stereo		**0.16** (± 0.24)	0.20 (± <u>0.25</u>)	**0.20** (± 0.23)	<u>0.26</u> (± <u>0.30</u>)	**0.08** (± 0.10)	0.70 (± **0.92**)
Passive Stereo		0.25 (± <u>0.34</u>)	**0.16** (± 0.17)	0.21 (± <u>0.25</u>)	0.25 (± 0.21)	0.17 (± 0.15)	1.26 (± <u>1.63</u>)
RF ToF	P^*	-0.15 (± **0.25**)	**-0.06** (± 0.21)	-0.10 (± **0.24**)	**-0.07** (± 0.17)	-0.07 (± **0.20**)	<u>+3.31</u> (± **1.53**)
NIR ToF		<u>+0.49</u> (± 0.43)	<u>+0.42</u> (± 0.44)	<u>+0.32</u> (± 0.57)	<u>+0.29</u> (± 0.39)	-0.30 (± 0.29)	+2.86 (± 1.67)
Active Stereo		**+0.06** (± <u>1.25</u>)	-0.07 (± <u>1.54</u>)	**-0.04** (± <u>1.67</u>)	+0.12 (± <u>1.09</u>)	**+0.04** (± 1.30)	**+0.82** (± <u>3.92</u>)
Passive Stereo		+0.22 (± 0.41)	+0.17 (± 0.38)	<u>+0.32</u> (± 0.85)	<u>+0.29</u> (± 0.48)	<u>+1.70</u> (± <u>8.95</u>)	+0.88 (± 2.49)
RF ToF	P_e^*	-0.14 (± 0.25)	**-0.06** (± 0.20)	**-0.10** (± 0.24)	**-0.07** (± **0.16**)	-0.07 (± <u>0.20</u>)	<u>+3.14</u> (± 1.41)
NIR ToF		<u>+0.51</u> (± 0.27)	<u>+0.38</u> (± 0.21)	<u>+0.28</u> (± **0.22**)	<u>+0.24</u> (± 0.17)	<u>-0.30</u> (± **0.09**)	+2.65 (± 1.29)
Active Stereo		**+0.00** (± **0.24**)	-0.13 (± <u>0.25</u>)	-0.14 (± 0.23)	+0.11 (± <u>0.30</u>)	**+0.00** (± 0.10)	**+0.45** (± **0.92**)
Passive Stereo		+0.05 (± <u>0.34</u>)	+0.11 (± **0.17**)	+0.15 (± <u>0.25</u>)	+0.23 (± 0.21)	+0.16 (± 0.15)	+1.23 (± <u>1.63</u>)

Table 11. We measure the depth deviation with respect to C_g, C_s, P, P_e and an additional signed version of P,P_e, which is denoted as P^*,P_e^*. All metrics are listed in the form ($\mu \pm \sigma$), consisting of the mean μ and standard deviation σ in centimeters, computed over the entire metric domain, respectively. The best results among all sensors of one metric type are highlighted in **bold** and the worst results are <u>underlined</u>.

.

	Metric Type	Mirror	Candle	Flowerpot (Transparent)	V1 Metal Plate	V2 Metal Plate	Hand Printed F
RF ToF		**0.87 (± 0.26)**	1.50 (± <u>1.12</u>)	1.31 (± <u>1.21</u>)	0.12 (± **0.05**)	**0.12 (± 0.05)**	<u>0.69 (± 0.86)</u>
NIR ToF	C_g	<u>3.77 (± 1.97)</u>	<u>2.04 (± 0.40)</u>	<u>2.73 (± 1.03)</u>	<u>0.77 (± 0.42)</u>	<u>0.74 (± 0.45)</u>	**0.12 (± 0.09)**
Active Stereo		2.13 (± 1.52)	**0.26 (± 0.29)**	**0.74 (± 0.53)**	**0.08 (± 0.06)**	0.30 (± 0.29)	0.17 (± 0.18)
Passive Stereo		2.31 (± 1.61)	1.64 (± 0.78)	2.01 (± 0.83)	0.13 (± 0.07)	0.15 (± 0.11)	0.20 (± 0.14)
RF ToF		**0.91 (± 0.14)**	<u>5.57 (± 2.78)</u>	1.86 (± 2.41)	0.13 (± **0.06**)	**0.13 (± 0.07)**	0.15 (± **0.10**)
NIR ToF	C_s	<u>33.31 (± 9.07)</u>	1.71 (± 0.49)	<u>3.10 (± 1.22)</u>	<u>0.81 (± 0.43)</u>	15.66 (± <u>17.32</u>)	**0.12 (± 0.10)**
Active Stereo		30.21 (± <u>14.59</u>)	**0.25 (± 0.26)**	**1.27 (± 1.78)**	**0.09 (± 0.07)**	5.54 (± 12.95)	0.21 (± <u>0.63</u>)
Passive Stereo		27.02 (± 11.33)	1.28 (± 0.65)	1.86 (± **0.93**)	0.16 (± 0.11)	0.20 (± 0.14)	<u>0.23 (± 0.16)</u>
RF ToF		**0.93 (± 0.12)**	<u>7.41 (± 3.79)</u>	2.74 (± 3.66)	0.11 (± 0.12)	**0.11 (± 0.12)**	**0.14 (± 0.20)**
NIR ToF	P	37.84 (± 14.84)	2.78 (± **0.35**)	<u>5.24 (± 2.04)</u>	<u>0.95 (± 0.48)</u>	16.23 (± <u>18.21</u>)	0.17 (± 0.30)
Active Stereo		<u>39.66 (± 24.75)</u>	**0.42 (± 0.49)**	**2.08 (± 2.30)**	**0.10 (± 0.13)**	6.16 (± 13.78)	<u>0.56 (± 2.02)</u>
Passive Stereo		30.82 (± 14.01)	2.10 (± 0.98)	3.50 (± **1.37**)	0.19 (± 0.15)	0.22 (± 0.20)	0.43 (± 0.73)
RF ToF		**0.93 (± 0.12)**	<u>7.37 (± 3.85)</u>	2.76 (± <u>3.66</u>)	0.10 (± 0.11)	**0.11 (± 0.12)**	**0.13 (± 0.20)**
NIR ToF	P_e	39.68 (± 6.57)	2.75 (± **0.15**)	<u>6.18 (± 1.79)</u>	<u>0.79 (± 0.39)</u>	24.92 (± <u>17.59</u>)	**0.13 (± 0.16)**
Active Stereo		<u>43.84 (± 20.28)</u>	**0.31 (± 0.44)**	**2.52 (± 1.77)**	**0.07 (± 0.09)**	7.46 (± 14.94)	<u>0.53 (± 1.07)</u>
Passive Stereo		35.96 (± 7.83)	2.15 (± 0.65)	4.36 (± **0.71**)	0.15 (± **0.09**)	0.23 (± 0.19)	0.34 (± 0.65)
RF ToF		**+0.93 (± 0.12)**	<u>+7.38 (± 3.79)</u>	+2.68 (± 3.66)	-0.06 (± 0.12)	**-0.06 (± 0.12)**	**-0.02 (± 0.20)**
NIR ToF	P^*	+37.84 (± 14.84)	+2.78 (± **0.35**)	<u>+5.24 (± 2.04)</u>	<u>+0.95 (± 0.48)</u>	+16.21 (± <u>18.21</u>)	<u>-0.11 (± 0.30)</u>
Active Stereo		<u>+39.58 (± 24.75)</u>	**+0.37 (± 0.49)**	**+2.00 (± 2.30)**	**+0.00 (± 0.13)**	+5.77 (± 13.78)	-0.03 (± <u>2.02</u>)
Passive Stereo		+30.80 (± 14.01)	+2.10 (± 0.98)	+3.49 (± **1.37**)	+0.17 (± 0.15)	+0.19 (± 0.20)	**+0.02 (± 0.73)**
RF ToF		**+0.93 (± 0.12)**	<u>+7.34 (± 3.85)</u>	+2.70 (± <u>3.66</u>)	-0.06 (± 0.11)	**-0.06 (± 0.12)**	**-0.01 (± 0.20)**
NIR ToF	P_e^*	+39.68 (± 6.57)	+2.75 (± **0.15**)	<u>+6.18 (± 1.79)</u>	<u>+0.79 (± 0.39)</u>	+24.90 (± <u>17.59</u>)	-0.11 (± **0.16**)
Active Stereo		<u>+43.84 (± 20.28)</u>	**+0.27 (± 0.44)**	**+2.52 (± 1.77)**	**+0.01 (± 0.09)**	+7.08 (± 14.94)	<u>-0.36 (± 1.07)</u>
Passive Stereo		+35.96 (± 7.83)	+2.15 (± 0.65)	+4.36 (± **0.71**)	+0.15 (± **0.09**)	+0.21 (± 0.19)	-0.07 (± 0.65)

	Metric Type	Hand Printed B	Hand Printed U	Metal Angle	Plunger	Silicone Cup	V1 Christmas Ball
RF ToF		<u>0.52 (± 0.70)</u>	<u>0.72 (± 0.80)</u>	0.47 (± 0.28)	0.62 (± 0.60)	<u>0.60 (± 0.47)</u>	0.59 (± 0.40)
NIR ToF	C_g	**0.09 (± 0.06)**	**0.12 (± 0.09)**	<u>0.90 (± 0.79)</u>	**0.23 (± 0.13)**	0.16 (± **0.10**)	0.22 (± 0.14)
Active Stereo		0.12 (± 0.11)	0.15 (± 0.12)	**0.23 (± 0.18)**	0.24 (± 0.17)	**0.13 (± 0.10)**	**0.13 (± 0.09)**
Passive Stereo		0.19 (± 0.16)	0.18 (± 0.14)	0.24 (± **0.17**)	<u>0.95 (± 0.64)</u>	0.15 (± **0.10**)	**0.13 (± 0.05)**
RF ToF		0.17 (± **0.12**)	0.18 (± **0.13**)	<u>1.20 (± 0.87)</u>	<u>1.18 (± 2.39)</u>	<u>1.43 (± 1.49)</u>	<u>0.75 (± 0.95)</u>
NIR ToF	C_s	**0.10 (± 0.12)**	**0.14 (± 0.15)**	1.02 (± 0.54)	**0.28 (± 0.19)**	0.21 (± 0.22)	0.41 (± 0.34)
Active Stereo		0.13 (± <u>0.25</u>)	0.17 (± <u>0.51</u>)	**0.29 (± 0.32)**	0.29 (± 0.32)	**0.13 (± 0.10)**	**0.14 (± 0.09)**
Passive Stereo		<u>0.20 (± 0.21)</u>	<u>0.22 (± 0.22)</u>	0.49 (± 0.57)	0.87 (± 0.63)	0.17 (± 0.14)	0.24 (± 0.15)
RF ToF		**0.17 (± 0.23)**	**0.19 (± 0.28)**	<u>1.71 (± 1.47)</u>	<u>1.47 (± 2.94)</u>	2.33 (± 2.79)	1.05 (± <u>1.41</u>)
NIR ToF	P	**0.17 (± 0.33)**	0.23 (± 0.45)	1.68 (± 1.29)	**0.42 (± 0.54)**	0.30 (± 0.38)	0.62 (± 0.64)
Active Stereo		0.26 (± <u>1.15</u>)	<u>0.46 (± 2.16)</u>	**0.49 (± 0.69)**	0.55 (± 0.96)	**0.20 (± 0.28)**	**0.22 (± 0.25)**
Passive Stereo		<u>0.33 (± 0.54)</u>	0.37 (± 0.62)	0.76 (± 0.88)	1.33 (± 1.35)	0.27 (± 0.37)	0.35 (± **0.22**)
RF ToF		0.17 (± 0.23)	0.19 (± 0.28)	<u>1.79 (± 1.47)</u>	1.62 (± <u>3.10</u>)	<u>2.55 (± 2.95)</u>	1.05 (± <u>1.41</u>)
NIR ToF	P_e	**0.08 (± 0.11)**	**0.16 (± 0.22)**	1.40 (± 0.76)	**0.27 (± 0.25)**	0.23 (± 0.27)	0.36 (± 0.37)
Active Stereo		0.18 (± 0.25)	0.28 (± 0.45)	**0.52 (± 0.74)**	0.58 (± 0.78)	**0.16 (± 0.16)**	**0.15 (± 0.19)**
Passive Stereo		<u>0.28 (± 0.40)</u>	<u>0.40 (± 0.58)</u>	0.67 (± **0.50**)	<u>1.71 (± 0.66)</u>	0.25 (± 0.21)	0.44 (± 0.23)
RF ToF		**-0.02 (± 0.23)**	**+0.00 (± 0.28)**	+1.56 (± <u>1.47</u>)	+0.89 (± <u>2.94</u>)	<u>+2.08 (± 2.79)</u>	<u>+0.96 (± 1.41)</u>
NIR ToF	P^*	**+0.02 (± 0.33)**	-0.01 (± 0.45)	<u>+1.66 (± 1.29)</u>	**+0.31 (± 0.54)**	+0.22 (± 0.38)	+0.50 (± 0.64)
Active Stereo		-0.03 (± <u>1.15</u>)	<u>+0.15 (± 2.16)</u>	**+0.40 (± 0.69)**	-0.32 (± 0.96)	**+0.05 (± 0.28)**	**+0.18 (± 0.25)**
Passive Stereo		<u>+0.08 (± 0.54)</u>	<u>+0.15 (± 0.62)</u>	+0.72 (± 0.88)	<u>-0.89 (± 1.35)</u>	+0.17 (± 0.37)	+0.34 (± **0.22**)
RF ToF		-0.01 (± 0.23)	**+0.00 (± 0.28)**	<u>+1.63 (± 1.47)</u>	+0.96 (± <u>3.10</u>)	<u>+2.27 (± 2.95)</u>	<u>+0.96 (± 1.41)</u>
NIR ToF	P_e^*	-0.02 (± **0.11**)	-0.04 (± **0.22**)	+1.37 (± 0.76)	**+0.26 (± 0.25)**	+0.19 (± 0.27)	+0.30 (± 0.37)
Active Stereo		<u>-0.08 (± 0.25)</u>	-0.02 (± 0.45)	**+0.39 (± 0.74)**	-0.55 (± 0.78)	**+0.12 (± 0.16)**	**+0.14 (± 0.19)**
Passive Stereo		**-0.00 (± 0.40)**	<u>+0.05 (± 0.58)</u>	+0.65 (± **0.50**)	<u>-1.71 (± 0.66)</u>	+0.22 (± 0.21)	+0.44 (± 0.23)

Table 12. We measure the depth deviation with respect to C_g, C_s, P, P_e and an additional signed version of P,P_e, which is denoted as P^*,P_e^*. All metrics are listed in the form ($\mu \pm \sigma$), consisting of the mean μ and standard deviation σ in centimeters, computed over the entire metric domain, respectively. The best results among all sensors of one metric type are highlighted in **bold** and the worst results are <u>underlined</u>.

	Metric Type	V2 Christmas Ball	V3 Christmas Ball	Water Cube	Flowerpot (Brown)	Brazen Rosette	Pool Ball
RF ToF	C_g	<u>0.59</u> (± <u>0.40</u>)	<u>0.60</u> (± <u>0.41</u>)	**0.16** (± **0.11**)	<u>1.00</u> (± <u>0.90</u>)	**0.11** (± **0.08**)	<u>1.28</u> (± <u>0.78</u>)
NIR ToF		**0.28** (± 0.18)	0.47 (± 0.24)	<u>3.00</u> (± <u>0.44</u>)	0.15 (± **0.08**)	<u>0.94</u> (± <u>0.35</u>)	0.64 (± **0.17**)
Active Stereo		0.51 (± 0.22)	0.50 (± 0.20)	0.56 (± 0.17)	**0.12** (± 0.21)	0.36 (± 0.24)	**0.31** (± 0.25)
Passive Stereo		0.46 (± **0.17**)	**0.30** (± **0.13**)	0.46 (± 0.24)	0.45 (± 0.25)	0.18 (± 0.13)	0.87 (± 0.30)
RF ToF	C_s	**0.10** (± **0.03**)	**0.10** (± **0.03**)	**0.12** (± **0.06**)	<u>0.53</u> (± <u>1.23</u>)	**0.11** (± **0.05**)	**0.09** (± **0.03**)
NIR ToF		<u>0.80</u> (± <u>0.80</u>)	<u>1.66</u> (± <u>1.65</u>)	<u>2.88</u> (± <u>0.54</u>)	0.22 (± 0.27)	<u>8.11</u> (± <u>11.87</u>)	0.69 (± <u>0.35</u>)
Active Stereo		0.43 (± 0.23)	0.39 (± 0.22)	0.54 (± 0.27)	**0.10** (± **0.09**)	0.46 (± 0.45)	0.33 (± 0.27)
Passive Stereo		0.46 (± 0.21)	0.39 (± 0.21)	0.52 (± 0.28)	0.49 (± 0.27)	0.29 (± 0.29)	<u>0.71</u> (± <u>0.35</u>)
RF ToF	P	**0.07** (± **0.06**)	**0.08** (± **0.08**)	**0.10** (± **0.10**)	<u>0.77</u> (± <u>1.96</u>)	**0.09** (± **0.13**)	**0.05** (± **0.07**)
NIR ToF		<u>1.05</u> (± <u>1.29</u>)	<u>3.19</u> (± <u>6.90</u>)	<u>3.99</u> (± <u>0.84</u>)	0.37 (± 0.46)	<u>9.03</u> (± <u>12.38</u>)	1.06 (± <u>0.48</u>)
Active Stereo		0.66 (± 0.24)	0.64 (± 0.25)	0.89 (± 0.35)	**0.16** (± **0.26**)	0.72 (± 0.82)	0.50 (± 0.38)
Passive Stereo		0.72 (± 0.22)	0.59 (± 0.23)	0.84 (± 0.35)	0.67 (± 0.37)	0.39 (± 0.48)	<u>1.15</u> (± 0.30)
RF ToF	P_e	**0.07** (± **0.06**)	**0.08** (± **0.08**)	**0.10** (± 0.10)	0.78 (± <u>1.97</u>)	**0.09** (± **0.13**)	**0.05** (± **0.07**)
NIR ToF		0.36 (± <u>0.57</u>)	<u>1.77</u> (± <u>2.11</u>)	<u>4.97</u> (± <u>0.27</u>)	0.15 (± **0.11**)	<u>15.00</u> (± <u>13.89</u>)	0.96 (± 0.18)
Active Stereo		0.75 (± 0.21)	0.73 (± 0.15)	0.86 (± 0.19)	**0.10** (± 0.12)	0.76 (± 0.78)	0.52 (± <u>0.37</u>)
Passive Stereo		<u>0.85</u> (± 0.16)	0.70 (± 0.18)	1.17 (± **0.05**)	<u>0.81</u> (± 0.25)	0.40 (± 0.44)	<u>1.29</u> (± 0.18)
RF ToF	P^*	**-0.05** (± **0.06**)	**-0.05** (± **0.08**)	**+0.08** (± **0.10**)	<u>+0.67</u> (± <u>1.96</u>)	**-0.02** (± **0.13**)	**+0.01** (± **0.07**)
NIR ToF		+0.68 (± <u>1.29</u>)	<u>+3.08</u> (± <u>6.90</u>)	<u>+3.99</u> (± <u>0.84</u>)	+0.36 (± 0.46)	<u>+9.01</u> (± <u>12.38</u>)	+1.05 (± <u>0.48</u>)
Active Stereo		+0.66 (± 0.24)	+0.63 (± 0.25)	+0.88 (± 0.35)	**-0.07** (± **0.26**)	-0.62 (± 0.82)	+0.48 (± 0.38)
Passive Stereo		<u>+0.72</u> (± 0.22)	+0.59 (± 0.23)	+0.83 (± 0.35)	+0.66 (± 0.37)	-0.23 (± 0.48)	<u>+1.15</u> (± 0.30)
RF ToF	P_e^*	**-0.05** (± **0.06**)	**-0.05** (± **0.08**)	**+0.08** (± 0.10)	+0.68 (± <u>1.97</u>)	**-0.02** (± **0.13**)	**+0.01** (± **0.07**)
NIR ToF		+0.17 (± <u>0.57</u>)	<u>+1.72</u> (± <u>2.11</u>)	<u>+4.97</u> (± <u>0.27</u>)	+0.15 (± **0.11**)	<u>+14.97</u> (± <u>13.89</u>)	+0.96 (± 0.18)
Active Stereo		+0.75 (± 0.21)	+0.73 (± 0.15)	+0.86 (± 0.19)	**-0.06** (± 0.12)	-0.70 (± 0.78)	+0.52 (± <u>0.37</u>)
Passive Stereo		<u>+0.85</u> (± 0.16)	+0.70 (± 0.18)	+1.17 (± **0.05**)	<u>+0.81</u> (± 0.25)	-0.30 (± 0.44)	<u>+1.29</u> (± 0.18)

	Metric Type	Polystyrene Plate	Bunny Box	Bunny
RF ToF	C_g	**2.10** (± 2.38)	**0.32** (± **0.28**)	<u>0.28</u> (± <u>0.22</u>)
NIR ToF		<u>3.28</u> (± <u>2.15</u>)	<u>0.50</u> (± 0.29)	0.26 (± 0.14)
Active Stereo		3.01 (± <u>3.05</u>)	0.39 (± 0.40)	0.12 (± 0.10)
Passive Stereo		2.99 (± 2.76)	0.39 (± <u>0.62</u>)	**0.08** (± **0.05**)
RF ToF	C_s	**0.14** (± **0.07**)	**0.30** (± **0.26**)	<u>0.50</u> (± <u>0.41</u>)
NIR ToF		2.50 (± 0.81)	0.48 (± 0.37)	0.31 (± 0.20)
Active Stereo		1.74 (± 1.97)	0.37 (± 0.35)	**0.12** (± **0.09**)
Passive Stereo		<u>2.92</u> (± <u>4.87</u>)	<u>0.61</u> (± <u>1.08</u>)	0.13 (± 0.13)
RF ToF	P	**0.12** (± **0.10**)	1.32 (± 2.45)	<u>0.74</u> (± 0.69)
NIR ToF		<u>17.95</u> (± <u>24.15</u>)	<u>1.91</u> (± <u>2.83</u>)	0.45 (± <u>1.77</u>)
Active Stereo		10.63 (± 15.21)	**0.99** (± **1.32**)	**0.16** (± **0.20**)
Passive Stereo		10.08 (± 11.66)	1.05 (± 1.88)	0.17 (± 0.27)
RF ToF	P_e	**0.12** (± **0.10**)	1.12 (± 2.15)	<u>0.74</u> (± <u>0.69</u>)
NIR ToF		<u>17.86</u> (± <u>24.12</u>)	**0.93** (± **0.87**)	0.38 (± **0.12**)
Active Stereo		10.48 (± 15.02)	1.46 (± 1.54)	0.11 (± **0.12**)
Passive Stereo		9.81 (± 11.41)	<u>2.68</u> (± <u>2.54</u>)	**0.10** (± **0.12**)
RF ToF	P^*	**+0.11** (± **0.10**)	+1.15 (± 2.45)	<u>+0.69</u> (± 0.69)
NIR ToF		<u>+17.95</u> (± <u>24.15</u>)	<u>+1.73</u> (± <u>2.83</u>)	-0.24 (± <u>1.77</u>)
Active Stereo		+10.63 (± 15.21)	-0.83 (± **1.32**)	-0.10 (± **0.20**)
Passive Stereo		+10.08 (± 11.66)	**-0.78** (± 1.88)	**+0.08** (± 0.27)
RF ToF	P_e^*	**+0.11** (± **0.10**)	+1.03 (± 2.15)	<u>+0.69</u> (± <u>0.69</u>)
NIR ToF		<u>+17.86</u> (± <u>24.12</u>)	**+0.92** (± **0.87**)	-0.38 (± **0.12**)
Active Stereo		+10.48 (± 15.02)	-1.46 (± 1.54)	-0.08 (± **0.12**)
Passive Stereo		+9.81 (± 11.41)	<u>-2.48</u> (± <u>2.54</u>)	**+0.07** (± **0.12**)

Table 13. We measure the depth deviation with respect to C_g, C_s, P, P_e and an additional signed version of P,P_e, which is denoted as P^*,P_e^*. All metrics are listed in the form ($\mu \pm \sigma$), consisting of the mean μ and standard deviation σ in centimeters, computed over the entire metric domain, respectively. The best results among all sensors of one metric type are highlighted in **bold** and the worst results are <u>underlined</u>.

.