

Automatic Spatial Calibration of Near-Field MIMO Radar With Respect to Optical Depth Sensors

Vanessa Wirth ^{✉,1} Johanna Bräunig² Danti Khouri² Florian Gutsche¹
 Martin Vossiek² Tim Weyrich^{*,1,3} Marc Stamminger^{*,1}

Abstract—Despite an emerging interest in MIMO radar, the utilization of its complementary strengths in combination with optical depth sensors has so far been limited to far-field applications, due to the challenges that arise from mutual sensor calibration in the near field. In fact, most related approaches in the autonomous industry propose target-based calibration methods using corner reflectors that have proven to be unsuitable for the near field. In contrast, we propose a novel, joint calibration approach for optical RGB-D sensors and MIMO radars that is designed to operate in the radar’s *near-field range*, within decimeters from the sensors. Our pipeline consists of a bespoke calibration target, allowing for automatic target detection and localization, followed by the spatial calibration of the two sensor coordinate systems through target registration. We validate our approach using two different depth sensing technologies from the optical domain. The experiments show the efficiency and accuracy of our calibration for various target displacements, as well as its robustness of our localization in terms of signal ambiguities.

I. INTRODUCTION

The ability to sense an environment in terms of accurate 3D information is crucial for many applications, including robotics, autonomous driving, or human-computer interaction. A prominent sensor class is range-sensing imagers; this work considers both optical imagers as well as imaging radar.

Driven by data availability, high spatial resolution, and low cost, optical depth sensing technologies such as time-of-flight cameras and single- or multi-view stereo algorithms are widely used; a tremendous amount of research has been conducted, for example, in the field of static [1] and dynamic [1]–[3] reconstruction, human pose and shape estimation [4], and scene understanding [5].

On the other hand, a growing interest has emerged with respect to radar imaging, prominently utilized for security scanning [6], [7] and autonomous driving [8], [9]. Radar is able to provide range cues in the presence of fog or dust, can penetrate fabric, and is insensitive to environmental light. Compared to camera-based systems, radar imaging is a recent range-sensing technology that involves calculating spatial object or feature distributions, commonly by using digital beamforming. Popular sensors are multiple-input multiple-output (MIMO) radars, which process

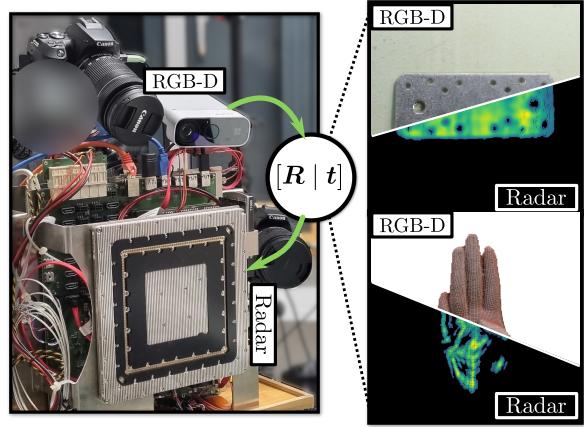


Fig. 1: Our calibration estimates the relative rotation R and translation t between an optical RGB-D sensor and an imaging MIMO radar incorporating high angular resolution.

the received signals coherently to form a synthetic antenna aperture by comparing the phase difference between multiple incoming signals at distinctive spatial receiver positions. They essentially exploit that each antenna (both transmitters and receivers) looks at scene points from different directions, which allows to 3D reconstruct a scene from the resulting phase differences. The result is commonly represented as a voxel grid or point cloud (cf. Figure 1), including confidence values about a target’s presence that are proportional to the received signal power.

A growing body of work [10]–[17] recognizes the potential of combining optical depth sensors (which we collectively refer to as *RGB-D sensors*), and MIMO radars. However, they invariably operate in the radar’s *far field*, at distances where standard solutions exist to mutually calibrate (align) the respective sensor coordinate systems. In contrast, we take on the unique challenge of localizing joint calibration objects within the MIMO radar’s *near-field range*, i.e., within few decimeters, where traditional radar targets appear strongly distorted.

While a small number of previous works in the context of autonomous driving compute the spatial calibration on the fly during the capture process, e.g., by leveraging motion cues [10], [18] during a car drive, most calibrations are static and target-based, that is, a specific calibration target [11]–[17] is designed to yield robust and accurate reconstruction results

* These authors contributed equally to this work

¹ Visual Computing Erlangen (VCE), Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

² Institute of Microwaves and Photonics (LHFT), Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

³ Dept of Computer Science, University College London, UK
 ✉vanessa.wirth@fau.de

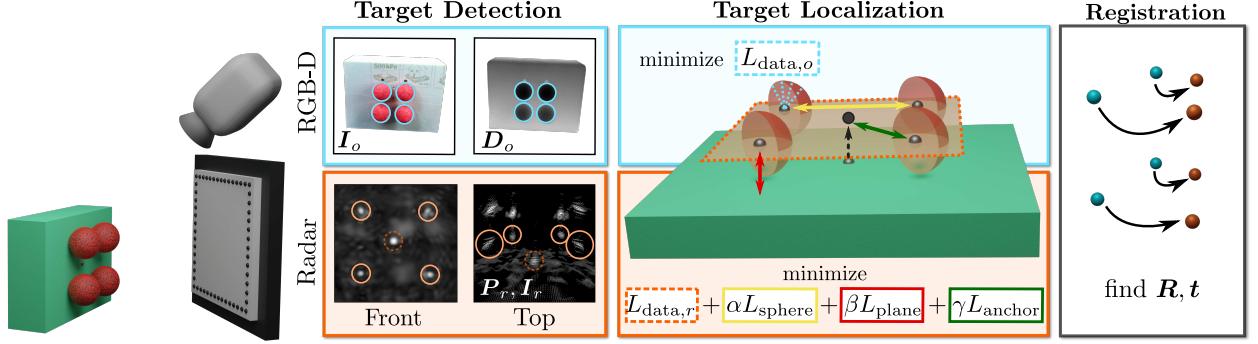


Fig. 2: The calibration is divided into sensor-specific parts for target detection and target localization. To acquire the calibration parameters, we register the localized target points from the optical domain (blue) to points of the radar domain (orange).

in all relevant sensors. The primary target of choice to be detected by a radar in the far field is a metal, trihedral corner reflector [12]–[17] because of its strong echo signal for a comparatively large range of acceptance angles: in far-field conditions, the signal propagation can be approximated as parallel to the radar’s line of sight, resulting in a retroreflective behavior for various antenna positions. Moreover, the reflector geometry ensures a total path length of the received signal that is constant across its entire aperture, which is why the corner is reconstructed as a bright, seemingly planar reflector that can easily be detected automatically.

For MIMO arrays in near-field scenarios with large angles between a target and a transmitter-receiver antenna pair, however, the desired properties of a corner reflector do not hold [19]. Figure 3 illustrates the distinctive signal response between a corner reflector captured in the far field and in the near field. For this reason, we conclude that the aforementioned line of related work is not suitable for calibration in the near field, where the target has only a few decimeters distance to the sensor. An orthogonal approach to ours proposed by Chen et al. [11] leverages optical markers as target to calibrate an imaging radar with a motion capture system for reconstruction of human bodies. The method introduced in this paper can be applied to optical RGB-D technologies with a 2D-3D correspondence relationship, for example time-of-flight and single- or multi-view stereo systems. These systems provide 2D depth maps, which can be back-projected into 3D given the camera’s intrinsic parameters. To the best of our knowledge, we are the first to propose an

automatic target-based calibration method for optical RGB-D sensors and imaging MIMO radars in the near field.

To achieve this, the contributions of this paper are:

- Design of a calibration target that is robustly detectable from various optical RGB-D sensors and MIMO radars.
- An automatic pipeline for target detection and localization, followed by a spatial registration.
- An overall framework that yields precise calibration parameters with millimeter accuracy, assessed by pairing a MIMO radar sensor with two different RGB-D technologies.

A. Overview

Our full pipeline is illustrated in Figure 2. We capture a calibration target, which is specifically designed for near-field conditions, from an optical RGB-D sensor and a MIMO radar. The optical sensor provides an RGB and a depth image, denoted as $I_o \in \mathbb{R}^{W \times H \times 3}$ and $D_o \in \mathbb{R}^{W \times H}$, respectively. Furthermore, we acquire a radar point cloud $P_r \in \mathbb{R}^{N \times 3}$ with confidence values proportional to the received signal amplitude $I_r \in [0, 1]^N$. Since the visibility of materials and geometries depends on the received signal wavelength, the target detection as well as the target localization are divided into sensor-specific parts. During detection, our method finds possible target candidates. Given these candidates, the localization stage utilizes sensor-specific prior knowledge about the calibration target to filter outliers and calculate the spatial position of the point samples that are used for the following registration stage. During registration, our method computes the optimal transformation between point samples from the optical and the radar domain, respectively. Lastly, for evaluation, we optionally employ an additional refinement stage with a second capture target.

II. CALIBRATION TARGET

To establish correspondences between an optical depth sensor and a MIMO radar, a calibration target is required that is robustly detectable, despite the significant domain gap between the different operating wavelengths. We opted for a target that can be detected within a wider range of viewing angles, to avoid having to precisely align the target in front the MIMO radar, as would be required for many potential target

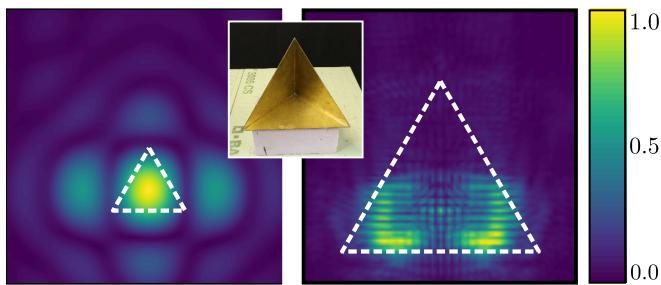


Fig. 3: Target confidence of a corner reflector captured by a MIMO radar at 2.6 m (left) and 0.3 m (right) distance.



Fig. 4: The calibration target consists of four styrofoam spheres ($\varnothing 5\text{ cm}$), each with a steel ball ($\varnothing 2.5\text{ mm}$) embedded at its center; sphere centers form a square of 6 cm edge length. A fifth steel ball is centered on the styrofoam back plane.

geometries where reconstruction quality significantly depends on the angle of incidence to the transmitters. Our calibration target is depicted in Figure 4 and consists of four textured styrofoam spheres, arranged in a square and mounted onto a styrofoam board. While styrofoam is a material hardly visible to radar, the spheres contain smaller, highly radar-reflecting steel balls inside, which were embedded using a high-precision drill. An additional steel ball at the center of the square is placed onto the styrofoam board. We will now elaborate on our design choices. First, we chose view-independent spherical shapes since hard corners and edges are challenging to detect in range sensors with millimeter accuracy due to multi-path signal interference at object silhouettes [20]. Furthermore, due to its material properties, metal is highly reflective for radar signals but sharp edges lead to diffraction, which introduces sidelobes and other types of noise into the reconstructed point cloud. Therefore, we took countermeasures to ensure the signal-to-noise ratio of our target is as high as possible. The square arrangement, which shares the symmetry and, approximately, the spatial extents of the MIMO antenna layout, provides a calibration feature distribution (see Section III) within the maximum focused area of the radar field of view, helps balancing out symmetric noise artifacts and enables the possibility of outlier detection through spatial constraints. Ensuring uniformity in the spot visible to each transmitting antenna, similar to the assumption made for corner reflectors in the far field, the diameter of the metal spheres is chosen such that the reconstructed signal is close to a single point target: in our setup, the diameter of 2.5 mm is smaller than the minimum transmitted wavelength. As the size of the spheres become very small this way, they become challenging to detect in RGB-D sensors simultaneously. For this reason, we embed the metal spheres that form a square at the center of comparably larger, 5 cm-diameter styrofoam spheres. To support optical stereo technologies, which rely on color features for high-quality depth reconstructions, we colored and textured the styrofoam spheres with random patterns. Note that, in this way, we induce noise into the radar reconstructions as well, since the material is not completely invisible anymore. In the next section, we elaborate on how to deal with this noise. Moreover, we ensure the spatial distance between both, metal and styrofoam spheres is sufficiently large to avoid multi-path effects. The final calibration target is placed in

front of both sensors such that the styrofoam spheres are inside their respective field of view.

III. AUTOMATIC TARGET DETECTION AND LOCALIZATION

The purpose of the target detection and localization stages is to automatically find the positions of the four metal balls, which are located at the center of the styrofoam spheres. Since optical and radar sensors operate on different wavelengths, the detection as well as the localization stage is sensor-specific. During target detection, we aim to find target candidates of point clusters in D_o and P_r , respectively. In the localization stage, we leverage photometric and spatial constraints to filter these candidates as well as to infer the metal ball locations.

A. Sensor-specific Target Detection

In the following, we describe the target detection for RGB-D sensors and MIMO radars separately.

1) *Target Detection in RGB-D Sensors:* Since the metal balls are not visible for RGB-D sensors, our method infers their spatial position from the styrofoam sphere centers. We detect the spheres by utilizing the 2D-3D correspondence of a depth map D_o and its respective RGB image I_o . Instead of detecting the sphere surface directly, we identify circles in the image plane. While spheres generally map to ellipses, for us this approximation still reliably detected the spheres. We utilize OpenCV's Circle Hough Transform [21] to find circles in D_o , in which we clamp the depth values to the near-field range of one meter. Note that, in case of stereo vision technologies, in which D_o is directly computed from I_o , this method can also be applied to RGB images instead. To summarize, our method produces a set of circle candidates, in which four of them are assumed to describe the projected surface of the styrofoam spheres in the image plane.

2) *Target Detection in MIMO Radars:* The metal balls inside the styrofoam spheres are highly reflective and appear as local maxima in the confidence values of the reconstructed point cloud P_r . Compared to the optical domain, reconstructions from radar imaging sensors are more prone to noise and automatic circle detection based on either a projected depth map or confidence map becomes challenging. In particular, we experienced scattered signals of possibly higher amplitude than the metal balls, which originate from the colored styrofoam spheres that are generally closer to the sensor (cf. radar top view in Figure 2). Contrary to related work, which locates corner reflectors based on the assumption of the brightest scatterer, i.e. the target of highest confidence value, we relax this assumption and detect multiple bright scatterers instead. Given P_r and I_r , we filter out noise and clutter with low confidence based on a decibel threshold t_{dB} . Next, our method uses greedy non-maximum suppression (GreedyNMS) [22] to find point clusters of high confidence with respect to a Euclidean distance threshold t_{min} . In other words, in each iteration we accept the point $p_r \in P_r$ of highest confidence as a point cluster and reject all other points belonging to the same cluster within a local neighborhood determined by t_{min} . Moreover, an additional threshold of a maximum Euclidean distance to all previously

selected clusters, t_{\max} , ensures to select further clusters close to previous ones and avoids the addition of signals from the background. After $N \geq 5$ cluster candidates are found, we iteratively assign points, ordered by confidence, to the nearest cluster, until each cluster has M samples. Based on these samples, we compute the centroid of each cluster. In this way, we acquire a set $C = \{\mathbf{c}^k \mid \mathbf{c}^k \in \mathbb{R}^3 \wedge (k = 1, \dots, N)\}$ of cluster centers, in which we assume that five of them belong to the metal balls of our calibration target.

B. Sensor-specific Target Localization

Analogously to the previous section, the localization of the metal ball centers (from amongst the candidates from the target detection stage) is sensor-specific and described in two separate sections.

1) Target Localization in RGB-D Sensors: Given the set of circle candidates from the detection stage, we utilize spatial and photometric constraints to find the four circles belonging to the styrofoam spheres. Similar to GreedyNMS, our method iterates through all detections that are ordered by circle confidence and filters duplicates as well as outliers based on two thresholds for color and size, respectively. The color threshold filters candidates on the basis of their median color deviation from the ground-truth. The size threshold discards candidates of significantly deviating radii from already selected circles. We continue the filtering procedure until four circles are acquired. Given the intrinsic parameters of \mathbf{D}_o , the circles are projected back into a 3D point cloud. To find the relative location of these point clouds with respect to their arrangement on the styrofoam board, we assume that the angle difference between the up-vector of the optical and radar sensor coordinate systems is less than 90° . Based on this assumption, we order the point clouds with respect to the up and right vector of the optical coordinate system. We denote the resulting set of sphere-shaped point clouds as $\mathcal{S}_o = \{S_o^j \mid S_o^j \in \mathbb{R}^{N^j \times 3} \wedge (j = 1, \dots, 4)\}$.

To locate the sphere centers $\Omega_o = (\mathbf{c}_o^1, \dots, \mathbf{c}_o^4)$ we minimize a weighted least-squares problem of a sphere equation $L_{\text{data},o}$ with known radius r that is fit to all points $s_o^j \in S_o^j$:

$$\Omega_o = \arg \min_{\tilde{\Omega}_o} L_{\text{data},o} \quad (1)$$

$$= \arg \min_{\tilde{\Omega}_o} \sum_i^4 \sum_j^{N^j} w_j (\|\tilde{\mathbf{c}}_o^i - s_o^j\|_2^2 - r^2). \quad (2)$$

Since optical depth sensors suffer from noise in particular at silhouettes, the point-wise error weight $w_j = \langle \mathbf{n}_j, \mathbf{e}_{\text{dir}} \rangle$ describes the range confidence of a point with normal \mathbf{n} with respect to the sensor's viewing direction \mathbf{e}_{dir} . To filter possible outliers, we minimize the energy term multiple times using RANSAC on the point clouds. During each iteration, we consider a random subset $\tilde{\mathcal{S}}_o^j$ with inlier ratio k and error e (from Equation 2) as the current best set $\tilde{\mathcal{S}}_{o,\text{best}}^j$ in case the following criterion is fulfilled:

$$\tilde{\mathcal{S}}_{o,\text{best}}^j = \begin{cases} \tilde{\mathcal{S}}_o^j & |k - k_{\text{best}}| > t_{\text{inl}} \vee e < e_{\text{best}}, \\ \tilde{\mathcal{S}}_{o,\text{best}}^j & \text{otherwise}. \end{cases} \quad (3)$$

An inlier ratio threshold, t_{inl} , offers a trade-off parameter between inlier maximization and error minimization.

2) Target Localization in MIMO Radars: To localize the five metal balls among the cluster centers detected in \mathcal{P}_r , we use their unique spatial topology established by design of the calibration target. In this way, it is possible to filter highly ambiguous candidates that may originate from the color-coated styrofoam spheres or other parts of the environment. Thereby, we utilize the fifth, central metal ball as an anchor point to localize the styrofoam board. Among all cluster centers in C , our method selects the best subset $\mathcal{S}_r = \{\mathbf{c}_r^k \mid \mathbf{c}_r^k \in C \wedge (k = 1, \dots, 4)\}$, together with the anchor point $\mathbf{c}_{a,r} \in C$, by minimizing the total weighted energy of:

$$S_r = \arg \min_{\tilde{\mathcal{S}}_r, \tilde{\mathbf{c}}_{r,a} \in C} L_{\text{data},r} + \alpha L_{\text{sphere}} + \beta L_{\text{plane}} + \gamma L_{\text{anchor}}. \quad (4)$$

Based on the a priori knowledge that four metal balls lie on a common plane, the term $L_{\text{data},r}$ minimizes the plane equation parameterized by normal $\mathbf{n} \in \mathbb{R}^3$ and reference point $\mathbf{k} \in \mathbb{R}^3$, of which its parameters are estimated along with \mathcal{S}_r :

$$L_{\text{data},r} = \sum_{\tilde{\mathbf{c}}_r \in \tilde{\mathcal{S}}_r} |\langle \tilde{\mathbf{c}}_r - \mathbf{k}, \mathbf{n} \rangle|. \quad (5)$$

The regularization term L_{sphere} enforces each sphere center pair $(\mathbf{c}^i, \mathbf{c}^j)$ to be close to the expected spatial distance $d^{i,j}$:

$$L_{\text{sphere}} = \sum_{i=1}^4 \sum_{j=i+1}^4 \left\| \mathbf{v}_{i,j} \cdot \left(\frac{1}{\|\mathbf{v}_{i,j}\|_2} - d^{i,j} \right) \right\|_2. \quad (6)$$

The vector $\mathbf{v}_{i,j} = f(\tilde{\mathbf{c}}_r^i) - f(\tilde{\mathbf{c}}_r^j)$ describes the relative distance between two spheres $\tilde{\mathbf{c}}_r^i, \tilde{\mathbf{c}}_r^j \in \tilde{\mathcal{S}}_r$, and the function $f(\mathbf{c}) = (\mathbf{c} - \mathbf{n} \cdot (\mathbf{c} - \mathbf{k}, \mathbf{n}))$ projects the center $\mathbf{c} \in \mathbb{R}^3$ onto the estimated plane. In a complementary manner, the term L_{plane} minimizes the distance d from a center to the styrofoam board, which is localized through the anchor point $\mathbf{c}_{a,r}$:

$$L_{\text{plane}} = \sum_{\tilde{\mathbf{c}}_r \in \tilde{\mathcal{S}}_r} |\langle \tilde{\mathbf{c}}_{a,r} - \tilde{\mathbf{c}}_r, \mathbf{n} \rangle - d|. \quad (7)$$

Lastly, the term L_{anchor} ensures that the anchor point $\mathbf{c}_{a,r}$ lies in the center of the square sphere arrangement:

$$L_{\text{anchor}} = \|f(\tilde{\mathbf{c}}_{a,r}) - \sum_{\tilde{\mathbf{c}}_r \in \tilde{\mathcal{S}}_r} \frac{1}{4} \tilde{\mathbf{c}}_r\|_2. \quad (8)$$

To find \mathcal{S}_r , our method tries $\binom{N}{5}$ combinations to sample five cluster centers $\tilde{\mathcal{S}}_r$ in each iteration. Since four of these centers should lie on a common plane, we test all possible $\binom{5}{4}$ combinations to find the plane (\mathbf{k}, \mathbf{n}) , which minimizes Equation 5. Similar to optical localization, the four metal ball candidates of the current sample are ordered with respect to the up and right vector of the sensor coordinate system. Next, the current sample is evaluated in terms of its inlier ratio, using Equation 5, and the error function given in Equation 4. Lastly, we compute Equation 3 to determine, whether the sample $\tilde{\mathcal{S}}_r$ is better than the current best random subset. To show the necessity of the additional spatial constraints in Equation 4, an ablation study is performed in Section V.

IV. AUTOMATIC SPATIAL REGISTRATION AND REFINEMENT

Based on the previously localized metal balls, the final stage automatically computes the relative rigid transformation between the coordinate systems of a sensor pair through spatial registration. Moreover, we support an optional refinement stage, utilizing a second object of simpler geometry to establish a significantly higher amount of correspondence pairs. As we will show, this optional stage confirms our method's accuracy but does not significantly improve results.

A. Calibration Parameter Estimation

Based on their spatial location on the styrofoam board, we find ordered pairs $(\mathbf{c}_r^i, \mathbf{c}_o^i)$ of sphere centers in the radar and optical sensor coordinate system, respectively. We assume that both sensor coordinate systems are metrical and use a priori knowledge about their factory settings to determine the uniform scale matrix $\mathbf{S} \in \mathbb{R}^{3 \times 3}$ from the optical to the radar coordinate units. Then, we solve for the optimal rotation $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ and translation $\mathbf{t} \in \mathbb{R}^3$ from the optical coordinate system to the radar coordinate system by minimizing their root mean square error. We use the closed-form solution of Kabsch [23] to acquire \mathbf{R} and \mathbf{t} as follows:

$$\mathbf{H} = \bar{\mathbf{C}}_r^T \cdot \mathbf{S} \cdot \bar{\mathbf{C}}_o \stackrel{\text{SVD}}{=} \mathbf{U} \cdot \Sigma \cdot \mathbf{V}^T \quad (9)$$

$$\mathbf{R} = \mathbf{U} \mathbf{V}^T \quad (10)$$

$$\mathbf{t} = \bar{\mathbf{c}}_r - \mathbf{R} \cdot \mathbf{S} \cdot \bar{\mathbf{c}}_o \quad (11)$$

The mean of the four sphere centers in sensor domain $*$ is denoted as $\bar{\mathbf{c}}_*$. The data matrices $\bar{\mathbf{C}}_r, \bar{\mathbf{C}}_o \in \mathbb{R}^{4 \times 3}$ contain the mean-centered spheres in pairwise order. Since \mathbf{S} is computed a priori, Σ is expected to be close to the identity matrix.

B. Calibration Refinement for Evaluation

As part of our evaluation (Section V), we employ an additional refinement stage as a method to assess potential calibration errors arising from uncertainties with respect to the sensor range. In this stage, the target for establishing correspondences is a simple, textured metal plate mounted on a styrofoam board for an upright standing. The amount of received signal from a planar surface is strongly view dependent for a MIMO radar such that the plate has to be placed parallel to the antenna aperture. Since it can be assumed that the first estimate of \mathbf{R} and \mathbf{t} is accurate enough, we compute correspondences via a projective mapping. Given the intrinsic parameters of \mathbf{D}_o , we transform all points in \mathbf{P}_r to the optical image plane and establish correspondence pairs based on points sharing the same pixel coordinate. Next, we repeatedly solve for \mathbf{R} and \mathbf{t} using the Kabsch algorithm in combination with RANSAC, in which we randomly sample correspondence pairs to minimize Equation 3.

V. EVALUATION

In this section, we describe the evaluation setup as well as the quality assessments of our calibration.

A. Evaluation Setup

The MIMO radar is a submodule of an Automotive Radome Tester provided by Rohde & Schwarz [24]. Its virtual aperture consists of 94×94 transmitting and receiving antennas, arranged on a square frame. The signal form is *stepped frequency continuous wave* within a frequency range from 72 GHz to 82 GHz with 128 frequency steps [25]. To acquire \mathbf{P}_r and \mathbf{I}_r , we make use of a state-of-the-art reconstruction method for millimeter wave imaging, which is known as back-projection and described further in [7], [25]. \mathbf{P}_r is reconstructed in a range between 20–65 cm. To demonstrate that our calibration target can be used for various optical technologies, we employ two distinct depth sensing methods: *amplitude modulated continuous wave* time-of-flight (ToF) using the Microsoft Kinect Azure [26] camera, and multi-view stereo (MVS) using five Canon DSLR cameras with > 24 MP resolution. The setup is depicted in Figure 5. We evaluate the calibration within a constrained environment using styrofoam as a rest table, a black screen made of fabric, and absorbers behind. Our experiments are divided into two scenarios: first, we record the calibration target, in a constrained position, followed by a more natural capture process. With respect to the former, we place the calibration target on a plastic turntable such that it is centered in the radar coordinate system. The anchor metal ball represents the center of its rotation. We place the turntable at 30 cm, 40 cm and 50 cm distance to the radar, respectively, and rotate the calibration target between $[-20^\circ, 20^\circ]$ in steps of 5° . To simulate a natural capture process, we relocate the calibration target multiple times within a range of 30–50 cm distance such that it is roughly centered by eye with respect to the radar. In this way, we acquire 40 different calibration target captures. Lastly, we record the refinement object once at 30 cm distance. We use this object in our experiments only if explicitly noted.

In our calibration pipeline, we set the parameters $t_{\text{dB}} = 15$, $t_{\min} = 2$ cm, $t_{\max} = 30$ cm, $t_{\text{inl}} = 0.05$, $N = 20$, $M = 7$, $\alpha = 2$, $\beta = 2$, and $\gamma = 4$. Furthermore, we set the maximum number of RANSAC iterations in the optical localization stage to 1000 and in the optional refinement stage to 100.

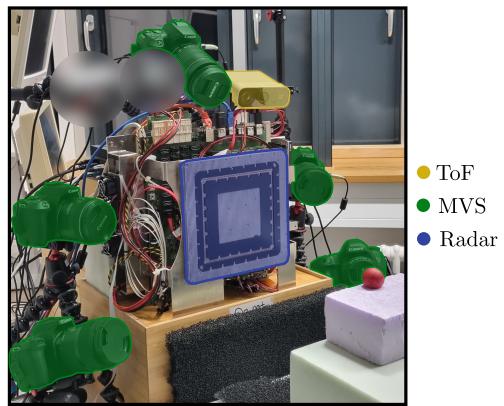


Fig. 5: Our setup consists of an imaging MIMO radar, a Kinect Azure camera (ToF) and five DSLR cameras (MVS).

B. Results

We show the efficiency of our method in three analyses. First, we evaluate the performance with respect to different orientations and distances of the calibration target. Second, we perform an ablation study to demonstrate the importance of the target design together with the utilization of spatial constraints in the radar domain. Lastly, we show qualitative results with respect to three captured objects. We captured objects of distinctive geometry and color, on which we measure the calibration error: a metal disk (at 30 cm and 40 cm distance to the MIMO radar), a symbol cut out from cardboard (at 30 cm), and a 3D-printed hand model coated in metal lacquer (at 30 cm). Our primary metric is the Chamfer distance C between object points $\mathbf{p}_o \in \mathbf{P}_o$ of an optical sensor and points $\mathbf{p}_r \in \mathbf{P}_r$ of the MIMO radar:

$$C = \frac{1}{2} \text{RMSE}(\mathbf{P}_o, \mathbf{P}_r) + \frac{1}{2} \text{RMSE}(\mathbf{P}_r, \mathbf{P}_o) \quad (12)$$

The root mean square error is calculated on the Euclidean norm per point pair, established based on nearest Euclidean distance. In the following, we will address each experiment in more detail.

1) View- and Distance-dependent Calibration: To demonstrate that our calibration target can be placed in front of a MIMO radar without giving a considerable amount of attention to its precise placement, we assess the calibration accuracy with respect to multiple orientation angles and distances. More specifically, we calculate the Chamfer distance of the cardboard symbol that was captured at 30 cm distance to the MIMO radar. Results are shown for both, the ToF-radar and MVS-radar sensor pair in Figure 6. Additionally, for both pairs, we ran the calibration 20 times for the same target at (30 cm, 0°), yielding a standard deviation of ±0.004° for the average rotation and ±0.17 mm for the average translation.

For both pairs, our method works best within all recorded orientation angles and a distance between 30–40 cm. Within this range, all samples of the ToF-radar and MVS-radar calibration have an average Chamfer distance of 1.69 mm and 1.72 mm, respectively, regardless of the target orientation. The average error increases by 0.53 mm and 0.54 mm

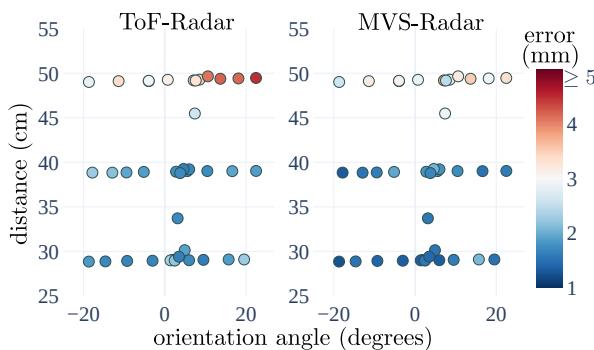


Fig. 6: The Chamfer distance for the ToF-radar and MVS-radar sensor pair, respectively. We plot the target angles and distances from the estimated plane in the radar coordinate system during sphere localization.

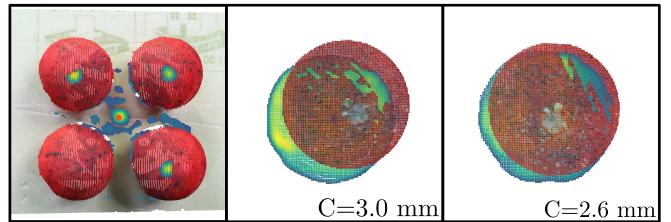


Fig. 7: MVS-radar calibration of a target at 50 cm distance (left). The anchor point (orange) approximately aligns with the radar signal. The error of a disk at 30 cm (middle) and 40 cm (right) decreases with its distance to the target.

when including the results at ≥50 cm distance. Upon further investigation, we observe that the Chamfer distance significantly depends on the spatial distance between the calibration location and the location of the evaluation object. In Figure 7, the results for a calibration target, captured at 50 cm distance, are depicted with respect to the metal disk, placed at 30 cm and 40 cm distance, respectively. This example illustrates that the error decreases with the distance between the calibration target location and the evaluation object. Hence, we conclude that the calibration is only valid within a specific range due to perspective distortion and systematic range errors of the RGB-D sensors. Moreover, the ToF camera exhibits a comparably large Chamfer distance at 50 cm distance and orientations > 5° when compared to the MVS-radar calibration. Assessed from the average calibration parameters, the ToF coordinate system has a spatial offset of {5 cm, 17 cm, 15 cm} and {12°, 15°, 20°} with respect to the (right, up, direction) vector triple of the radar coordinate system. As a consequence, the angle between the target plane and the ToF view direction is +16° larger than in the radar coordinate system, such that results at 22° in Figure 6 have an orientation of 38° in the ToF coordinate system. To summarize, the application-dependent working distance has to be considered during target calibration. Inside this working distance, our approach achieves millimeter accuracy far below the radar wavelength (3.7 mm) and the random noise distribution (≤ 17 mm) of the ToF camera. Since our calibration stays accurate, regardless of the target orientation in our experiments, we conclude that it is not required to balance the target precisely in front of the sensors.

2) Ablation Study: We demonstrate the necessity of the choices made during calibration target design as well as the utilization of these choices through spatial constraints in an ablation study. The results with respect to the Chamfer distance of the cardboard symbol are given in Figure 9. Since our main contribution lies in a near-field calibration of an imaging MIMO radar, the focus of this study is on target localization in the radar domain. Without any systematic spatial arrangement of the spheres, the only term that can be applied during localization is $L_{\text{data},r}$ and the resulting calibration error is within 5 cm on average. By arranging the spheres in a square, the additional regularization term L_{sphere} can be employed, which decreases the calibration error by

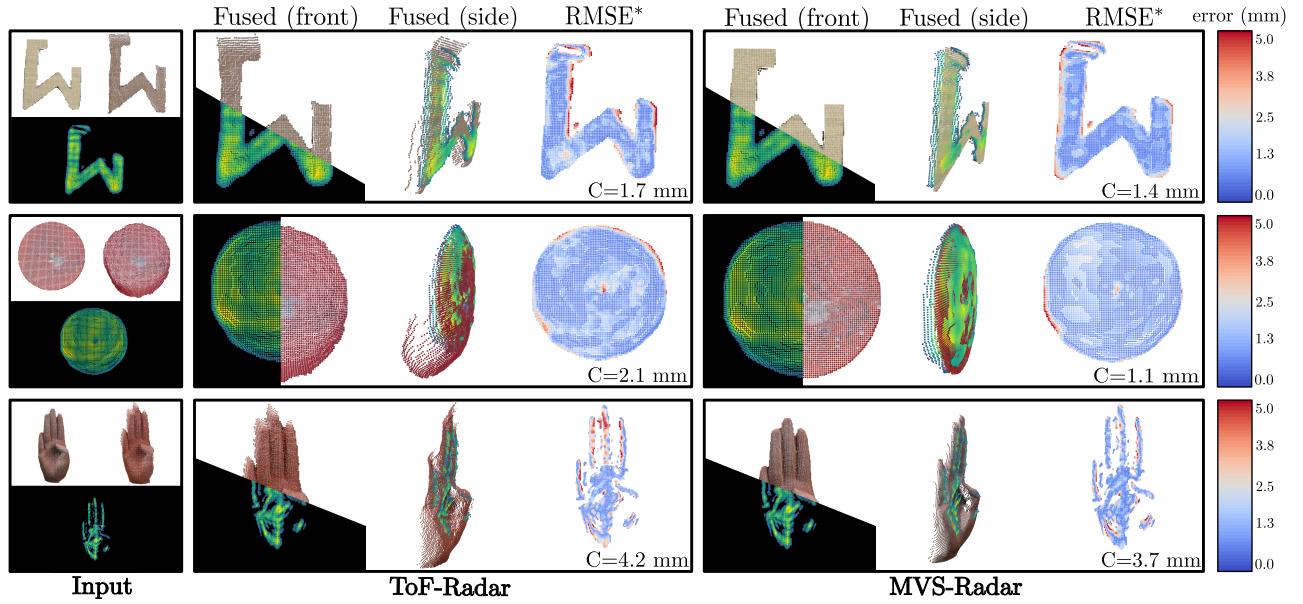


Fig. 8: We show qualitative results for a cardboard symbol, a disk, and a 3D printed hand, respectively. The Chamfer distance C is indicated in the bottom of the error visualization. RMSE* denotes the point-wise RMSE between a radar point and the nearest point of Euclidean distance from the optical sensor.

4 cm on average. Lastly, the anchor point that is mounted in the center of the square on the styrofoam board leads to another error decrease by 8 mm through the utilization of the regularization terms L_{plane} and L_{anchor} . To conclude, the results demonstrate that the square arrangement with five metal balls is necessary to achieve a calibration quality below 2 mm. So far, our calibration is single-shot, which means it only requires one capture of the calibration target and, thus, little capture effort. We further show results of a second capture, in which we record the metal plate and perform the refinement stage. While the median of both, the ToF-radar (1.86 mm) and MVS-radar (1.72 mm) calibrations are similar to calibration without refinement (1.92 mm and 1.72 mm), we observe a common decrease in the error variance, together with the mean. This decrease is due to the

improving calibration results specifically at ≥ 45 cm distance, since the refinement target is placed at 30 cm and, therefore, is able to correct the misalignment arising from the spatial distance between the position of calibration target and the cardboard symbol. However, we argue that in these cases it would have been simpler to place the calibration target at 30 cm in the first place. In summary, we demonstrate that each of our target design choices is necessary and the calibration can not be further improved by a second capture.

3) *Qualitative Results:* In the last experiment, the calibration accuracy of a target placed at 30 cm distance is assessed in qualitative results using the three additional captured objects of distinctive geometry complexity and distance. In Figure 8, we estimate the Chamfer distance and the point-wise RMSE from a radar point to the nearest point, and show results for the cardboard, the metal disk and the hand, each recorded at a distance of 30 cm. For all objects, the RMSE is primarily below 2 mm. Its distribution is geometry- and sensor-specific. Moreover, the ToF-radar alignment results in higher point-wise errors due to the fact that active ToF cameras exhibit more noise than high-resolution passive MVS algorithms. The cardboard symbol in the first row has the best average alignment quality for both, ToF-radar and MVS-radar calibrations. The disk in the second row exhibits a comparably higher Chamfer distance due to flying pixel artifacts in the ToF camera. Lastly, the hand in the third row of Figure 8 demonstrates the huge domain gap between reconstructions of optical and radar sensors. For complex geometries, most of the signals do not return back to the MIMO radar, which underlines the importance of careful calibration target design. In summary, the qualitative results further corroborate the accuracy of our calibration method and offer interesting findings in terms of the domain-specific sensor characteristics.

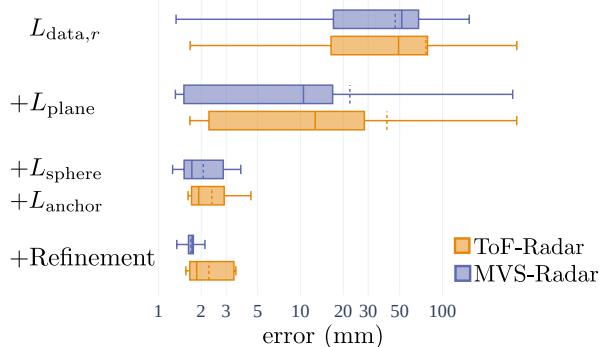


Fig. 9: We vary the spatial constraints during radar sphere localization to simulate the availability of target elements added during the design. We also assess the calibration quality after the refinement stage. The median and the mean of each box plot are marked as solid and dashed lines, respectively.

VI. CONCLUSION

We presented a novel calibration method for optical technologies in combination with an imaging MIMO radar in the near field within centimeter range. Considering the large domain gap between the two frequency domains, we designed a suitable calibration target that consists of four textured styrofoam and five metal balls, arranged at the corners and the center of a square. Given a capture of this target, our method detects circles in the optical domain, and clusters points of high target confidence in the radar domain. Due to careful design of the target's spatial arrangement, we utilized photometric as well as spatial constraints to detect and localize the four metal balls within each sensor coordinate system. Finally, we compute the calibration parameters through spatial registration of these balls and propose to assess the alignment quality in an optional refinement stage. In the evaluation, we demonstrate the effectiveness of our target design, eliminating the need for careful positioning in front of a sensor, and show the importance of the spatial arrangement. In summary, our calibration target is single-capture, user-friendly with respect to its placement, and yields millimeter accuracy up to a remaining error of less than 2 mm, which is considerably small such that it may originate from sensor noise itself.

ACKNOWLEDGEMENT

The authors would like to thank the Rohde & Schwarz GmbH & Co. KG (Munich, Germany) for providing the radar imaging devices. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1483 – Project-ID 442419336, EmpkinS.

REFERENCES

- [1] M. Zollhöfer, P. Stotko, A. Görlitz, *et al.*, “State of the art on 3d reconstruction with rgb-d cameras,” *Computer Graphics Forum*, vol. 37, no. 2, pp. 625–652, 2018. doi: <https://doi.org/10.1111/cgf.13386>.
- [2] X.-F. Han, H. Laga, and M. Bennamoun, “Image-based 3d object reconstruction: State-of-the-art and trends in the deep learning era,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1578–1604, 2021. doi: [10.1109/TPAMI.2019.2954885](https://doi.org/10.1109/TPAMI.2019.2954885).
- [3] E. Tretschk, N. Kairanda, M. B R, *et al.*, “State of the art in dense monocular non-rigid 3d reconstruction,” *Computer Graphics Forum*, vol. 42, no. 2, pp. 485–520, 2023. doi: <https://doi.org/10.1111/cgf.14774>.
- [4] Y. Tian, H. Zhang, Y. Liu, and L. Wang, “Recovering 3d human mesh from monocular images: A survey.” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 45, no. 12, pp. 15 406–15 425, 2023, ISSN: 1939-3539. doi: [10.1109/TPAMI.2023.3298850](https://doi.org/10.1109/TPAMI.2023.3298850).
- [5] M. Naseer, S. Khan, and F. Porikli, “Indoor scene understanding in 2.5/3d for autonomous agents: A survey,” *IEEE Access*, vol. 7, pp. 1859–1887, 2019. doi: [10.1109/ACCESS.2018.2886133](https://doi.org/10.1109/ACCESS.2018.2886133).
- [6] S. S. Ahmed and L.-P. Schmidt, “Illumination of humans in active millimeter-wave multistatic imaging,” in *2012 6th European Conference on Antennas and Propagation (EUCAP)*, 2012, pp. 1755–1757. doi: [10.1109/EuCAP.2012.6206694](https://doi.org/10.1109/EuCAP.2012.6206694).
- [7] S. S. Ahmed, “Microwave imaging in security — two decades of innovation,” *IEEE Journal of Microwaves*, vol. 1, no. 1, pp. 191–201, 2021. doi: [10.1109/JMW.2020.3035790](https://doi.org/10.1109/JMW.2020.3035790).
- [8] D. Schwarz, N. Riese, I. Dorsch, and C. Waldschmidt, “System performance of a 79 ghz high-resolution 4d imaging mimo radar with 1728 virtual channels,” *IEEE Journal of Microwaves*, vol. 2, no. 4, pp. 637–647, 2022. doi: [10.1109/JMW.2022.3196454](https://doi.org/10.1109/JMW.2022.3196454).
- [9] O. Bialer, A. Jonas, and T. Tirer, “Super resolution wide aperture automotive radar,” *IEEE Sensors Journal*, vol. 21, no. 16, pp. 17 846–17 858, 2021. doi: [10.1109/JSEN.2021.3085677](https://doi.org/10.1109/JSEN.2021.3085677).
- [10] L. Heng, “Automatic targetless extrinsic calibration of multiple 3d lidars and radars,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 10 669–10 675. doi: [10.1109/IROS45743.2020.9340866](https://doi.org/10.1109/IROS45743.2020.9340866).
- [11] A. Chen, X. Wang, S. Zhu, Y. Li, J. Chen, and Q. Ye, “Mmbody benchmark: 3d body reconstruction dataset and analysis for millimeter wave radar,” ser. MM ’22, Association for Computing Machinery, 2022, 3501–3510, ISBN: 9781450392037. doi: [10.1145/3503161.3548262](https://doi.org/10.1145/3503161.3548262).
- [12] J. Domhof, J. F. Kooij, and D. M. Gavrila, “An extrinsic calibration tool for radar, camera and lidar,” in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 8107–8113. doi: [10.1109/ICRA.2019.8794186](https://doi.org/10.1109/ICRA.2019.8794186).
- [13] C.-L. Lee, Y.-H. Hsueh, C.-C. Wang, and W.-C. Lin, “Extrinsic and temporal calibration of automotive radar and 3d lidar,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 9976–9983. doi: [10.1109/IROS45743.2020.9341715](https://doi.org/10.1109/IROS45743.2020.9341715).
- [14] L. Cheng, A. Sengupta, and S. Cao, “3d radar and camera co-calibration: A flexible and accurate method for target-based extrinsic calibration,” in *2023 IEEE Radar Conference (RadarConf23)*, 2023, pp. 1–6. doi: [10.1109/RadarConf2351548](https://doi.org/10.1109/RadarConf2351548).
- [15] C.-L. Lee, C.-Y. Hou, C.-C. Wang, and W.-C. Lin, “Extrinsic and temporal calibration of automotive radar and 3-d lidar in factory and on-road calibration settings,” *IEEE Open Journal of Intelligent Transportation Systems*, vol. 4, pp. 708–719, 2023. doi: [10.1109/OJITS.2023.3312660](https://doi.org/10.1109/OJITS.2023.3312660).
- [16] S. Agrawal, S. Bhandari, K. Doycheva, and G. Elger, “Static multitarget-based autocalibration of rgb cameras, 3-d radar, and 3-d lidar sensors,” *IEEE Sensors Journal*, vol. 23, no. 18, pp. 21 493–21 505, 2023. doi: [10.1109/JSEN.2023.3300957](https://doi.org/10.1109/JSEN.2023.3300957).
- [17] M. Choi, S. Yang, S. Han, *et al.*, “Msc-rad4r: Ros-based automotive dataset with 4d radar,” *IEEE Robotics and Automation Letters*, vol. 8, no. 11, pp. 7194–7201, 2023. doi: [10.1109/LRA.2023.3307005](https://doi.org/10.1109/LRA.2023.3307005).
- [18] E. Wise, Q. Cheng, and J. Kelly, “Spatiotemporal calibration of 3-d millimetre-wavelength radar-camera pairs,” *IEEE Transactions on Robotics*, vol. 39, no. 6, pp. 4552–4566, 2023. doi: [10.1109/TRO.2023.3311680](https://doi.org/10.1109/TRO.2023.3311680).
- [19] Y. Liu and X. Xu, “Mimo radar images of a trihedral corner reflector for near-field measurement,” in *2017 International Applied Computational Electromagnetics Society Symposium (ACES)*, 2017, pp. 1–2.
- [20] P. Zanuttigh, L. Minto, G. Marin, F. Dominio, and G. Cortelazzo, “Time-of-flight and structured light depth cameras: Technology and applications.” Jan. 2016, pp. 1–355, ISBN: 978-3-319-30971-2. doi: [10.1007/978-3-319-30973-6](https://doi.org/10.1007/978-3-319-30973-6).
- [21] *The opencv reference manual: Hough circle transform*, OpenCV, 2023. [Online]. Available: https://docs.opencv.org/4.9.0/d4/d70/tutorial_hough_circle.html.
- [22] J. Hosang, R. Benenson, and B. Schiele, “Learning non-maximum suppression,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6469–6477. doi: [10.1109/CVPR.2017.685](https://doi.org/10.1109/CVPR.2017.685).
- [23] W. Kabsch, “A solution for the best rotation to relate two sets of vectors,” *Acta Crystallographica Section A*, vol. 32, no. 5, pp. 922–923, 1976. doi: [10.1107/S0567739476001873](https://doi.org/10.1107/S0567739476001873). [Online]. Available: <https://doi.org/10.1107/S0567739476001873>.
- [24] *R&s@qar50 quality automotive radome tester*, Rohde&Schwarz, 2023. [Online]. Available: https://www.rohde-schwarz.com/products/test-and-measurement/radome-tester/rs-qar50-quality-automotive-radome-tester_63493-1138625.html?change_c=true.
- [25] J. Bräunig, V. Wirth, C. Kammel, *et al.*, “An ultra-efficient approach for high-resolution mimo radar imaging of human hand poses,” *IEEE Transactions on Radar Systems*, vol. 1, pp. 468–480, 2023. doi: [10.1109/TRS.2023.3309574](https://doi.org/10.1109/TRS.2023.3309574).
- [26] *Azure kinect dk hardware specifications*, Microsoft, 2022. [Online]. Available: <https://learn.microsoft.com/en-us/azure/kinect-dk/hardware-specification>.