

Low-cost Extrinsic Calibration Between LIDAR and Thermal Camera for Indoor Mapping

Benjamin Ronald van Manen^{1,2}, Ville Lehtola², Abeje Yenehun Mersha¹, Francesco Nex²

¹ Academy of Life Science, Engineering And Design, Saxion University of Applied Sciences, Enschede, The Netherlands
- (b.r.vanmanen, a.y.mersha)@saxion.nl

² Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands
- (v.v.lehtola, f.nex)@utwente.nl

Keywords: Extrinsic calibration, Thermal camera, LIDAR, UGV, UXV, Mapping

Abstract

Building on the concept of LIDAR–RGB extrinsic calibration using simple box targets (Pusztai and Hajder, 2017), we propose a low-cost and effective method for extrinsic calibration between a thermal camera and a LIDAR. Our approach utilizes a single, inverted half-box target featuring a heated asymmetric circles grid on one of its planes, designed for high visibility in thermal imagery. Thermal camera poses are estimated using OpenCV’s calibration toolbox, while LIDAR poses are computed by extracting the three planar surfaces of the calibration target. The extrinsic transformation between the thermal camera and the LIDAR is then determined using a cycle consistency constraint. We evaluate the method in terms of calibration accuracy and precision. Additionally, the setup can be used to assess the time synchronization between the thermal and LIDAR data streams by visualizing the projection errors on dynamic scene elements. Additional results can be found in the video at youtu.be/00DHZe4rVec. The source code is available at github.com/SaxionMechatronics/CycleCameraLiDARCalibration.

1. Introduction

Multi-sensor systems play a crucial role in a wide range of applications, including uncrewed systems and vehicles (UXS/UXV), by providing complementary, multi-modal data. To enable reliable data fusion for downstream tasks, both intrinsic and extrinsic calibration of the sensors is essential (Yeong et al., 2021). Our research is motivated by firefighting scenarios in smoke-filled environments, where an Unmanned Ground Vehicle (UGV) equipped with a LiDAR and a thermal camera is used to create 3D maps of the fire incident to increase the situational awareness of firefighters. The thermal imaging and LiDAR pointclouds are combined to see through the smoke to capture the surrounding geometry. While this is the primary use case, the proposed calibration method is general and can be applied across different platforms, sensors and applications.

The essential idea in the proposed thermal-LiDAR extrinsic calibration method is to leverage the strength of each sensor: *planar surfaces* serve as reliable features for LiDAR, while a *pattern target*, designed to be clearly visible in thermal imagery, provides robust features for the camera. This work is inspired by the use of ordinary boxes to perform a low-cost LIDAR-RGB extrinsic calibration (Pusztai and Hajder, 2017). However, unlike RGB cameras, thermal cameras capture temperature differences rather than visual features, making standard boxes often indistinct in thermal imagery. To address this, we introduce a temperature gradient using heating elements to form a clearly visible asymmetric circles grid in the thermal images. This pattern is embedded on one face of an inverted box structure composed of three planar surfaces.

Our contributions are:

1. The design of a low-cost calibration target for thermal-LiDAR extrinsic calibration with clear plane boundaries and guaranteeing a visible pattern at all times.

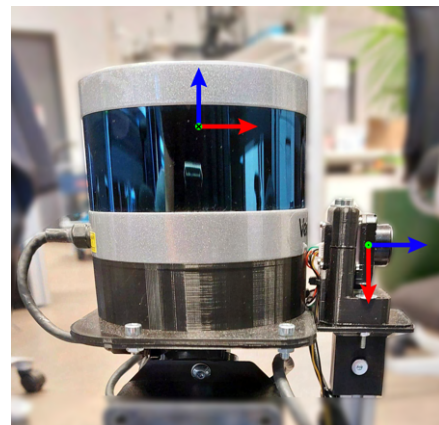


Figure 1. Sensor setup to be calibrated. A Velodyne VLP16 LiDAR with 16 channels and a FLIR Boson+ Industrial thermal camera with a 95° horizontal field-of-view. The thermal camera is placed a few centimeters to the left of the LiDAR’s x-axis.

The x-, y-, and z-axis are denoted by the red, green and blue arrows respectively. Following the right hand rule, the thermal camera’s x-axis is pointing towards the reader while the LiDAR’s y-axis is pointing away.

2. The method for a suitable thermal-LiDAR calibration procedure requiring a single target and no training, where we form a cycle consistency constraint between the camera and the LiDAR from two distinct poses.
3. The calibration results to show the precision and accuracy, followed by elaborative discussion.

Our sensor setup is displayed in Fig. 1. The paper is structured as follows. Related work is reviewed in section 2, after which we propose our calibration approach in section 3. Section 4 presents the evaluation of the results and a discussion of the their limitations. Finally, this paper is concluded in section 5.

2. Related work

Based on the surveyed literature, thermal-LiDAR extrinsic calibration methods may be categorized into two categories: target-based techniques and targetless techniques. Target-based calibration approaches rely on a known physical pattern and/or surfaces to reliably extract robust features in the different modalities. Targetless methods extract features from the natural scene, making them more flexible and faster to implement. The quality of the features is however dependent on the environment.

Target-based techniques rely on identifying corresponding features in both camera and LiDAR data; however, this is particularly challenging with thermal cameras due to low resolution and limited texture, and with LiDAR due to its sparse, textureless nature, making feature matching across modalities difficult. Using a well-designed calibration target can help mitigate these issues by providing distinct, high-contrast features that are reliably detectable in both sensor modalities. (Borrmann et al., 2012) track the movement of a planar target to estimate the extrinsic transformation between a LiDAR and a thermal camera. The planar target is embedded with a symmetric circular pattern created by 30 tiny incandescent light bulbs to compute the pose of the target with respect to the thermal camera. The light bulbs are detected by first creating a binary image based on a temperature threshold. After applying a second binary filter that removes blobs that are too small or too large, the final mask should contain 30 distinct blobs. The planar target is detected in the LiDAR pointcloud using RANSAC plane extraction. The location of the symmetric light bulb circular pattern is then matched on the extracted 3D plane using Iterative Closest Point (ICP) to obtain the 3D coordinates of the light bulb pattern. The extrinsic transformation between the LiDAR and the thermal camera is finally found by minimizing the reprojection error of the light bulb pattern from LiDAR to thermal image. Similar to (Borrmann et al., 2012), (Yuan et al., 2022) uses a symmetric circular pattern as a target. To make the pattern visible in the thermal images. The circles grid is however cut out of black carbon fiber. Behind the carbon fiber plate is a thin heated sheet of aluminium. The authors employ the Maximum Stable Extremal Regions (MSER) and arc-support ellipse fitting to detect the pattern circles rather than OpenCV's *findCirclesGrid()* as they state that MSER is robust to the image noise in the low-resolution thermal images. The target plane is extracted from the LiDAR data using RANSAC and edge detection based on depth discontinuities between the front carbon fiber plate and the back aluminium plate. As in (Borrmann et al., 2012), the 3D location of the holes is then projected onto the detected target plane and the extrinsic parameters are estimated by minimizing the reprojection error from 3D LiDAR to 2D thermal camera. (Zhang et al., 2018) demonstrates a two-phase target-based approach to obtain the extrinsic transformation between a LiDAR and a thermal camera by first estimating the extrinsic parameters between the LiDAR and the RGB camera, and then from the RGB camera to the thermal camera. The extrinsic transformation from the LiDAR to the RGB camera is obtained from the KITTI Calibration Toolbox which is provided with time-synchronized observations in both modalities of a collection of large chessboard patterns at different orientations inside a empty room. Through stereo calibration using an asymmetric circles grid, the extrinsic transformation from the RGB camera to the thermal camera is obtained. The calibration pattern visible in both image modalities is constructed using a heated plate with a laser-cut asymmetric circles grid. The final extrinsic calibration from LiDAR to thermal camera is finally equal to the

transformation from LiDAR to RGB camera times the transformation from RGB camera to thermal camera. (Zhang et al., 2019) employs a heated plate with four large holes as a calibration target to estimate the extrinsic parameters between a sparse LiDAR and a low-resolution thermal image. This approach is an adaptation of (Guindel et al., 2017) for LiDAR-stereo RGB extrinsic calibration. The 3D coordinates of the centers of the four circles are first extracted as in (Guindel et al., 2017). The 3D coordinates of the four centers in the thermal camera frame are obtained by first detecting the circles in a binary mask of the thermal image. The homography matrix mapping the 3D circle centers in world coordinates to 2D coordinates in the thermal image is estimated. The obtained homography matrix is then used to calculate the 3D coordinates of the centers in the thermal camera frame. Finally, the extrinsic parameters between the LiDAR and the thermal camera are estimated by matching the pairs of four 3D circle centers as in (Guindel et al., 2017). (Dalirani et al., 2023) created a large square calibration target for RGB-thermal-LiDAR extrinsic calibration. The target has a checkerboard pattern including Aruco markers. Each corner of the checkerboard is equipped with thermal resistors to make them visible in the thermal images. The four corners of the target plane are also marked with brighter thermal resistors. The locations of all the resistors are obtained using blob detection. The edges of the target plane are obtained using RANSAC line equation estimation and the location of the detected bright resistors. Using RANSAC again, the target plane and its four edges are extracted from the 3D LiDAR data. Corresponding lines and plane are then matched in both modalities. The initial extrinsic rotation and translation are obtained by Singular Value Decomposition (SVD) and by solving a linear system respectively. The extrinsic parameters are further refined using non-linear optimization by minimizing the point-to-plane and point-to-line reprojection. The authors state that due to the size of the calibration target, a single pair of LiDAR and thermal data is sufficient to obtain a correct estimation of the extrinsic transformation. The accuracy can be further increased by using multiple different poses.

Targetless approaches for thermal-LiDAR extrinsic calibration have also been proposed. They rely on ground truth training data (Mharolkar et al., 2022) or on environments with high thermal contrasts (Yang et al., 2024, Fu et al., 2022). In (Mharolkar et al., 2022), the estimation of the extrinsic transformation between a LiDAR and RGB camera, and between a LiDAR and a thermal camera is performed online using a single pair of LiDAR scan and thermal image through a deep-learning network. The LiDAR scan is projected into a 2D image using the thermal camera's intrinsic matrix. Features are then extracted from both modalities with ResNet-34. Features are then matched in a separate block of consecutive convolutional layers of increasing number of channels. The extrinsic rotation (quaternion) and translation are finally estimated in parallel in two regression blocks. The loss function used for training is the weighted linear combination of the translational and rotation error between the estimated and ground truth extrinsic parameters. The approach is trained and validated on the KITTI360 dataset and the NTU Dataset using the provided extrinsic parameters as ground truth. (Fu et al., 2022) proposes a targetless calibration approach to estimate the extrinsic transformations between a LiDAR, a thermal camera and a stereo RGB camera. First SIFT features are extracted and matched in the stereo pair to obtain a sparse 3D pointcloud from the RGB stereo camera. The extrinsic transformation between the LiDAR and the stereo RGB camera are obtained by aligned both pointclouds through

Multiframe Iterative Closest Point (MFICP) algorithm. To obtain features to estimate the extrinsic transformation between the LiDAR and the thermal camera, edges are first extracted in the thermal images using the Canny filter and in the LiDAR pointcloud using depth discontinuities across scanning lines. The extrinsic transformation is then estimated by Reprojection Edge Alignment Error (REAE), i.e. by minimizing the distance of the points on the 3D edges projected to their closest edge in the thermal image. (Yang et al., 2024) proposes a targetless LiDAR-RGB and LiDAR-thermal calibration that focuses on extracting and matching edges across the different modalities. LiDAR edges are obtained by computing the intersection lines of neighboring planes extracted from a dense pointcloud. The dense pointcloud is obtained by accumulating and matching scans from a mechanical LiDAR using the LiDAR-inertial odometry algorithm Fast-LIO. Edges are extracted from the thermal images by combining the first- and second-order derivatives of the image. Edges from both modalities are then matched together through k-Nearest Neighbors (kNN) matching by projecting the LiDAR edges into the thermal image using an initial guess of the extrinsic transformation between the sensors. The extrinsic transformation is then refined using Perspective n-Lines optimization. The approach is validated both in simulations and real-world data, using a single pair of LiDAR-thermal data and multiple consecutive pairs.

When comparing various target-based thermal-LiDAR extrinsic calibration methods, a common approach in (Yuan et al., 2022, Zhang et al., 2018, Zhang et al., 2019) involves preheating the element that produces contrast in the thermal image. However, during data acquisition, this heated element gradually cools down, causing the thermal pattern to become less distinct with respect to the background over time. Additionally, all target-based methods reviewed used a single target plane, with which it is often challenging, especially with low-resolution LiDARs like the Velodyne VLP16, to consistently capture all four edges of the target surface in the pointcloud while simultaneously covering the entire thermal image frame. Achieving accurate extrinsic calibration that generalizes across the full frame during point cloud projection requires a large calibration target as in (Dalirani et al., 2023) if only a single data pair is used for optimization. This, in turn, increases both the cost and complexity of the calibration setup. The targetless approaches proposed by (Fu et al., 2022, Yang et al., 2024) depend on naturally occurring thermal gradients in the environment to extract robust edge features that can also be detected by the LiDAR. However, in practice, such conditions are difficult to achieve, particularly in environments at room temperature. Furthermore, supervised learning methods as in (Mharolkar et al., 2022) that aim to estimate the extrinsic transformation in real-time require ground truth extrinsic values for training. These methods essentially attempt to replicate, on-the-fly, the results typically obtained through prior target-based calibration. This raises important questions about the ability of supervised learning approaches to generalize to previously unseen environments and to different sensor configurations.

3. Calibration approach

The proposed approach, described in Fig. 2, relies on tracking a calibration target in both the thermal and LiDAR modalities. First the asymmetric circles grid is detected in the thermal images to obtain the thermal camera's poses with respect to the world frame. Secondly, the LiDAR world poses are estimated

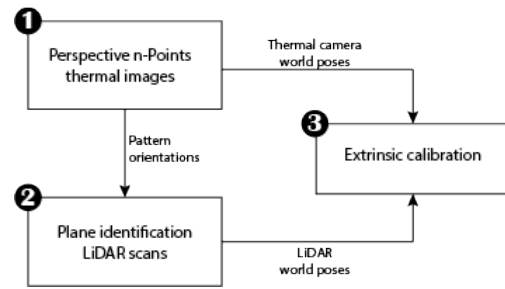


Figure 2. Flowchart of the proposed thermal-LiDAR extrinsic calibration approach

by detecting the three planes of the calibration target in the scans. Here, the orientation of the calibration pattern in the corresponding thermal images is used to help identify the three planes. Lastly, corresponding thermal-LiDAR measurements are randomly matched in pairs to compute relative pose transformations and estimate the extrinsic transformation between the two sensors by imposing a cycle consistency constraint. The sensor setup to be calibrated including a FLIR Boson+ Industrial thermal camera and a Velodyne VLP16 LiDAR is depicted in Fig. 1.

The design of the calibration target is shown in Fig. 3. It is an inverted half-box with one wall equipped with a heated asymmetric circles grid as in Fig. 4b visible by the thermal camera. We have used low-cost elements such as multiplex wooden plates, foam and aluminium. A bang-bang controller with a single thermo-couple in the center circle assures the calibration pattern maintains high contrast through the entire calibration process. The mat black coated aluminium pucks inserted in the circles act as a low-pass filter preserving the heat as the heating elements are switched on and off by the controller.

3.1 Optimization problem

Notation. Transformations between 6 degree-of-freedom poses are expressed as 4×4 homogeneous transformation matrices, as follows:

- $T_L^{(i)}$: The pose of the LiDAR in the world frame at timestamp i expressed as a homogeneous transformation from the LiDAR coordinate system L to some fixed world frame W .
- $T_C^{(i)}$: The pose of the thermal camera in the world frame at timestamp i expressed as a homogeneous transformation

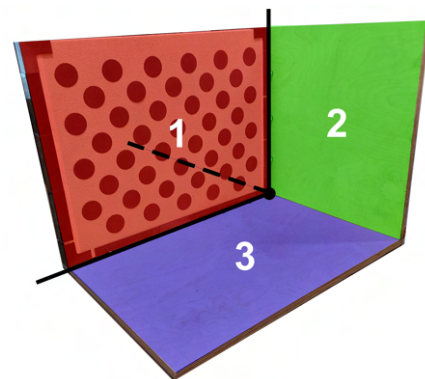


Figure 3. Calibration target plane description and coordinate system (following right-hand rule)

from the thermal camera coordinate system C to the same world frame W .

- $T_L^{(ij)}$: The relative transformation from the pose of the LiDAR at timestamp i to the pose of the LiDAR at timestamp j , defined as $T_L^{(ij)} = (T_L^{(j)})^{-1} T_L^{(i)}$.
- $T_C^{(ij)}$: The relative transformation from the pose of the thermal camera at timestamp i to the pose of the thermal camera at timestamp j , defined as $T_C^{(ij)} = (T_C^{(j)})^{-1} T_C^{(i)}$.
- T_L^C : The extrinsic transformation, the homogeneous transformation mapping from the LiDAR frame L to the thermal camera frame C .

Cycle consistency constraint. We form the following product of transformations, which ideally should equal to the identity matrix I :

$$\Delta_{(ij)} = (T_L^{(ij)})^{-1} (T_L^C)^{-1} T_C^{(ij)} T_L^C. \quad (1)$$

This cycle is interpreted from right to left as:

1. Start in LiDAR frame L at timestamp i ,
2. move to camera frame C at timestamp i ,
3. then to camera frame C at timestamp j ,
4. then to LiDAR frame L at timestamp j ,
5. and back to LiDAR frame L at timestamp i .

If T_L^C is perfectly consistent with a *single* rigid extrinsic transformation between LiDAR and camera, then $\Delta_{(ij)} = I$. We assume that the mount is rigid, so $T_L^C = \text{const} \quad \forall (i, j)$. Thus, it is noise or imperfect estimates that may cause $\Delta_{(ij)} \neq I$.

Least-squares objective. The cycle constraint is computed over random pairs of timestamps (i, j) . Enforcing the cycle consistency constraint over randomly-assigned pairs instead of consecutive timestamps ensures that there is sufficient difference in the two sensor poses. This is important to guarantee successful convergence towards the correct solution as the relative poses are then further from identity matrix.

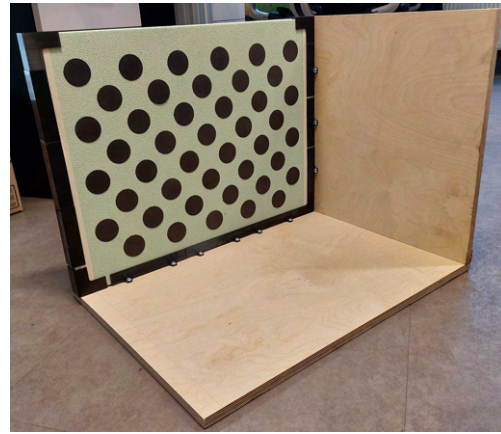
To enforce consistency across *all* pairs (i, j) , we penalize the deviation of each $\Delta_{(ij)}$ of Eq. 1 from I . A standard choice is to use the matrix logarithm in $SE(3)$, mapping $\Delta_{(ij)}$ to a 6D residual vector. Hence, we seek the extrinsic transformation T_L^C by optimizing

$$\min_{T_L^C} \sum_{(ij)} \|\text{Log}(\Delta_{(ij)})\|^2. \quad (2)$$

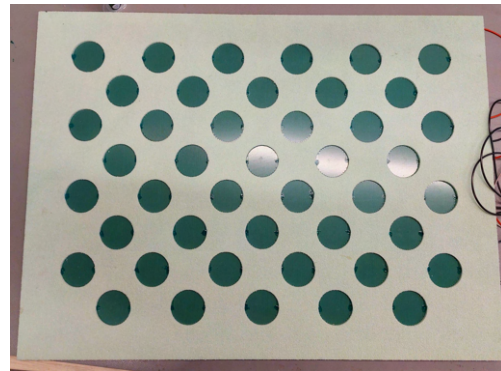
In practice, $\text{Log}(\Delta_{(ij)}) \in se(3) \simeq R^6$ in Eq. (2) measures how far $\Delta_{(ij)}$ is from the identity, thus ensuring the “cycle closure” is satisfied in an optimal least-squares sense.

Available transformations. From the thermal camera tracking of the heated asymmetric circles grid and Perspective n-Points (PnP), we have estimates of the relative transformations $T_C^{(ij)}$, which describe the camera’s motion from the pose at timestamp i to the pose at timestamp j . From the LiDAR plane-fitting (and

intersection point estimation), we likewise obtain estimates of the relative transformations $T_L^{(ij)}$, describing the LiDAR’s motion between the poses of the same pair (i, j) of timestamps. Both sets of transformations are derived in isolation, and hence may not be mutually consistent.



(a)



(b)

Figure 4. (a) Calibration target used for the estimation of the extrinsic parameters between the LiDAR and the thermal camera. The target is comprised of three orthogonal planes, where the front plane has a heated asymmetric circles grid visible by the thermal camera. (b) The mat black coated aluminium pucks of the asymmetric circles grid are heated using temperature-controlled heating elements.

The extrinsic transformation is defined by six parameters which are estimated by the optimization problem: three translations components and three Rodrigues rotation vector components (Valdenebro, 2016). The non-linear problem is solved using the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm (Liu and Nocedal, 1989) using *Scipy*’s *minimize* function and using *JAX* (Bradbury et al., 2018) to compute the cost function gradient.

3.2 Relative transformation calculation of the thermal camera

The relative transformations $T_C^{(ij)}$ between pairs of thermal camera world poses is computed as described in Algorithm 1. For processing convenience and software compatibility (OpenCV), the 16-bit raw thermal data is linearly mapped to an 8-bit range to generate the thermal images, preserving only the dynamic range from the 16-bit histogram. Then a binary mask is generated, where the threshold is manually chosen such that the circles are clearly distinguished from the foam around

Algorithm 1 Thermal camera relative transformation estimation using the asymmetric circles grid

```

1: Input: Sequence of 8-bit thermal images  $\{I_1, I_2, \dots, I_N\}$ , camera matrix  $K_t$ , distortion coefficients  $d_t$ 
2: Output: Camera poses  $\{\mathbf{T}_C^{(i)}\}$ , relative transformations  $\{\mathbf{T}_C^{(ij)}\}$ 
3: for each image  $I_i$  in the sequence do
4:    $\text{Mask}_i \leftarrow \text{BinaryThreshold}(I_i)$ 
5:    $\text{pattern\_detected} \leftarrow \text{DetectCircularGrid}(\text{Mask}_i)$ 
6:   if  $\text{pattern\_detected}$  then
7:      $\text{points2D}_i \leftarrow \text{SavePatternPoints}(I_i)$ 
8:   end if
9: end for
10: for each  $\text{points2D}_i$  in detected pattern points do
11:    $\mathbf{T}_C^{(i)} \leftarrow \text{SolvePnP}(\text{points2D}_i, \text{pattern3D})$ 
12: end for
13: for each pair  $i, j$  in pairs do
14:    $\mathbf{T}_C^{(ij)} \leftarrow (\mathbf{T}_C^{(j)})^{-1} \cdot \mathbf{T}_C^{(i)}$ 
15: end for

```

the calibration pattern. The calibration pattern is then detected in the obtained binary mask using OpenCV's *findCirclesGrid()* function.

The pose of the thermal camera with respect to the calibration target is then computed using Perspective n-Points (PnP), where the 3D coordinates of each circle in the pattern is expressed following the world coordinate system defined in Fig. 3.

The thermal camera's intrinsic parameters were obtained prior to the thermal-LiDAR calibration by utilizing the same calibration board and pattern detection method as explained above. The intrinsic parameters were estimated using OpenCV's *calibrateCamera()* function.

3.3 Relative transformation calculation of the LiDAR

The relative transformations $\mathbf{T}_L^{(ij)}$ between LiDAR scans are obtained as described in Algorithm 2.

Planes are initially detected in the LiDAR scan using iterative RANSAC plane fitting. We assume that the three planes of the calibration target are always the closest planes to the LiDAR sensor. Due to the nature of iterative RANSAC, points near the intersections of planes are sometimes incorrectly assigned at the surface intersections. To mitigate this, the detected plane boundaries are refined by reassigning each point associated with the calibration target to its nearest plane, based on the obtained plane equations.

The three calibration planes are then identified using their normal vectors and the orientation of the calibration pattern, as detected in the corresponding thermal image. The normal vectors are adjusted to ensure they point toward the interior of the calibration structure (i.e., toward each other). We first assume that the pattern has an angle $< 45^\circ$ in the corresponding thermal image. In that case, the normal vector of plane 3 is pointing upwards and therefore has the highest z-component of the three. Plane 1 and plane 2 are distinguished based on their normal vectors' y-components: under the assumption that the calibration target is oriented such that all three planes are visible to the LiDAR, the y-component of the normal vector of plane 1 is always less than or equal to zero, while that of plane 2 is greater than or equal to zero. If the pattern angle in the thermal image is $\geq 45^\circ$, the plane indices are cyclically shifted downward by one (i.e., plane 1 becomes plane 3, plane 2 becomes plane 1, etc.). Outliers (due to wrongly detected planes) are automatically removed by only accepting scans where the three detected planes have a minimum number of points, the orthogonality between

the three intersection lines is within a certain threshold and the obtained intersection point of the lines is within a certain distance from the mean location of the three planes.

Having identified the three planes, the intersection line $\hat{\mathbf{n}}_{12}$ between plane 1 and plane 2 as well as the intersection line $\hat{\mathbf{n}}_{13}$ between plane 1 and plane 3 are computed using the plane equations. The intersection point P , also known as the origin of the world coordinate frame, is obtained by finding the intersection $\hat{\mathbf{n}}_{12} \cap \hat{\mathbf{n}}_{13}$. The orientation of the LiDAR in world coordinates is then defined as in Eq. 3 following the right hand rule:

$$\mathbf{R}_L^W = [\hat{\mathbf{n}}_{12} \times \hat{\mathbf{n}}_{13}, \hat{\mathbf{n}}_{13}, \hat{\mathbf{n}}_{12}]^T \quad (3)$$

where both $\hat{\mathbf{n}}_{12}$ and $\hat{\mathbf{n}}_{13}$ are 3×1 vectors. The LiDAR pose $\mathbf{T}_L^{(i)}$ is then defined as in Eq. 4:

$$\mathbf{T}_L^{(i)} = \begin{bmatrix} \mathbf{R}_L^W & -\mathbf{R}_L^W P \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \quad (4)$$

Algorithm 2 LiDAR relative transformation calculations via plane intersection detection

```

Input: Sequence of LiDAR scans  $\{S_1, S_2, \dots, S_N\}$ 
2: Output: Poses  $\{\mathbf{T}_L^{(i)}\}$ , relative transforms  $\{\mathbf{T}_L^{(ij)}\}$ 
for each scan  $S_i$  do
3:    $\mathcal{P}_i \leftarrow \text{ExtractPlanes}(S_i)$ 
4:    $\mathcal{P}_i^{\text{near}} \leftarrow \text{FindThreeClosestPlanes}(\mathcal{P}_i)$ 
5:    $\mathcal{P}_i^{\text{refined}} \leftarrow \text{RefinePlaneBoundaries}(\mathcal{P}_i^{\text{near}})$ 
6:    $\mathcal{P}_i^{\text{identified}} \leftarrow \text{IdentifyPlanes}(\mathcal{P}_i^{\text{refined}}, \text{Fig. 3})$ 
7:    $\mathcal{L}_i \leftarrow \text{ComputePlaneIntersections}(\mathcal{P}_i^{\text{identified}})$ 
8:    $\mathbf{x}_i \leftarrow \text{ComputeIntersectionPoint}(\mathcal{P}_i^{\text{identified}})$ 
9:    $\mathbf{T}_L^{(i)} \leftarrow \text{ComputePoseFromPlanes}(\mathcal{P}_i^{\text{identified}}, \mathcal{L}_i, \mathbf{x}_i)$ 
10: end for
11: for each pair  $i, j$  in pairs do
12:    $\mathbf{T}_L^{(ij)} \leftarrow (\mathbf{T}_L^{(j)})^{-1} \cdot \mathbf{T}_L^{(i)}$ 
13: end for

```

4. Results

In our test setup, there are about $N = 1003$ inlier poses over which the estimation is done. The number of poses is high, because the LiDAR captures data at 10 Hz frequency. The camera has 30 Hz frequency, and the closest image in time is used for matching against the LiDAR scan. The time difference between the LiDAR scans and their matched thermal images is 8.5 ± 4.8 milliseconds. The calibration dataset is acquired by holding the calibration target at different distances up to five meters and different orientations in front of the sensor setup. Each posture of the target is held still for about half a second before moving to the next one. The entire dataset was recorded uninterrupted and lasted less than ten minutes.

Orthogonality of the planes is estimated by computing the dot product between the normal vectors of the LiDAR-measured planes, and it is of the order of $1e^{-12}$. Even if the plane normals are not orthogonal, an orthonormal basis can be calculated using the normals, and the thermal pattern can be set to span e.g. the yz-plane of that basis.

Detection examples of the calibration target are displayed in Fig. 5. The difficulty in obtaining robust detections lies in covering the entire thermal frame over time while also always ensuring that all three planes are visible in the LiDAR scans and

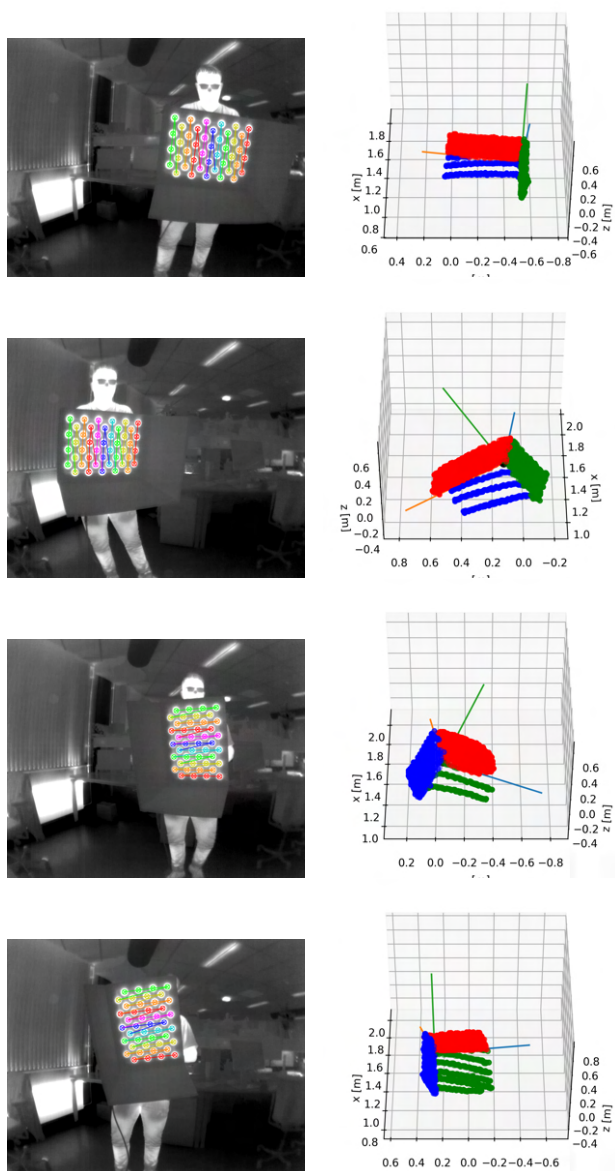


Figure 5. Detection examples of the calibration target in the thermal camera (left) and in the corresponding LiDAR pointcloud (right). In the two upper rows, the pattern angle is $< 45^\circ$ while it is $\geq 45^\circ$ in the bottom two rows.

that the calibration pattern is distinguishable in the thermal images. Being too close to the sensors, the planes are partially outside of the LiDAR's field-of-view or inside its dead-zone and cannot be detected correctly. When being further away to have more vertical freedom in the LiDAR, the risk of not being able to detect the pattern as it becomes one big white blob in the thermal image increases. Our tests showed that remaining between 1.5 and 3.5 meters from the sensor setup is optimal. Furthermore, preventing plane 1 with the calibration pattern from being parallel with the thermal camera plane guarantees that the corner of the inverted half-box is visible in the LiDAR scans. The control temperature of the asymmetric circles grid is set at 28°C . This temperature produced good contrast to reliably detect the asymmetric circles grid across the thermal images while limiting heat transfer to the surrounding foam.

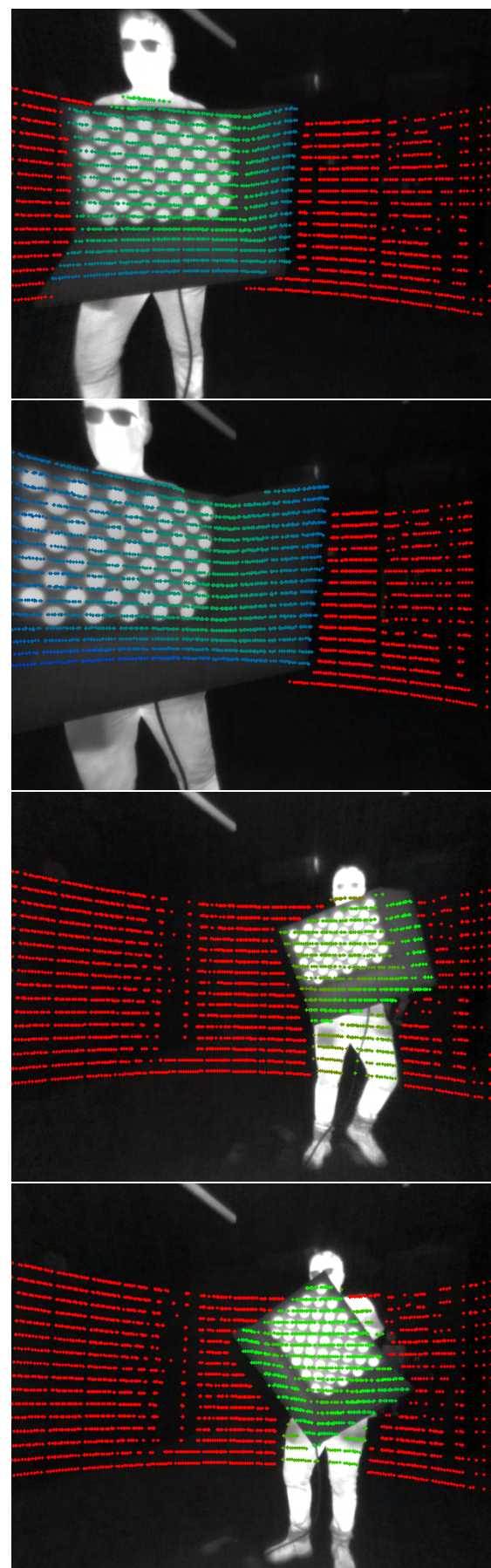


Figure 6. Examples of LiDAR projections onto thermal images using the estimated extrinsic calibration parameters.

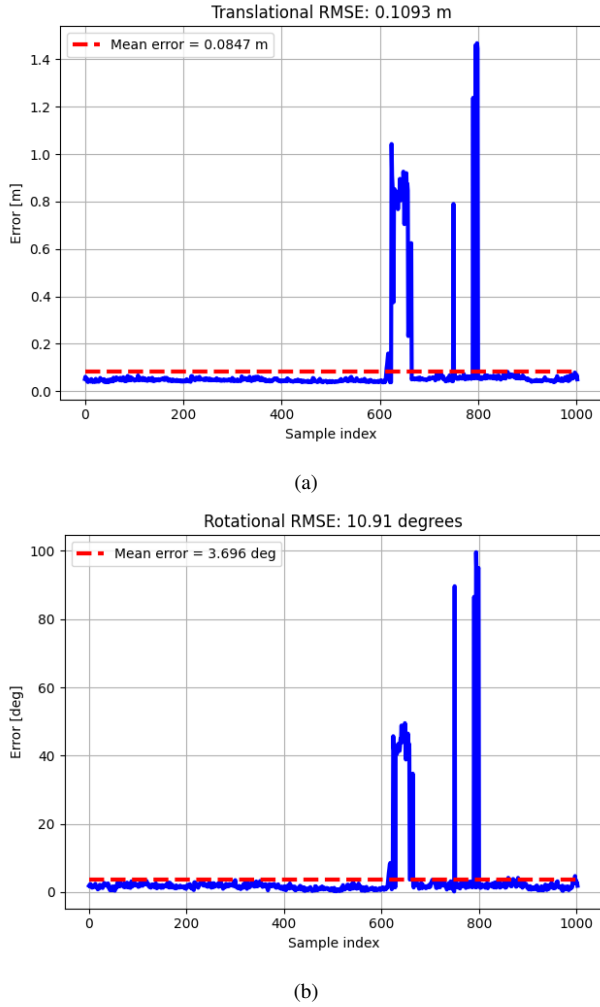


Figure 7. (a) Translational and (b) rotational error between $\hat{T}_L^{(i)}$ and $T_L^{(i)}$ using the final extrinsic parameters on the entire calibration dataset.

The extrinsic transformation $\hat{T}_L^C$ obtained for the sensor setup in Fig. 1 is presented in Eq. (5) and standard error $\sigma_{\hat{T}_L^C}$ in Eq. (6):

$$\hat{T}_L^C = \begin{bmatrix} 0.0396 & -0.9987 & 0.0307 & 0.0367 \\ 0.0386 & -0.0292 & -0.9988 & -0.0716 \\ 0.9985 & 0.0407 & 0.0374 & -0.0834 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \quad (5)$$

$$\sigma_{\hat{T}_L^C} = \begin{bmatrix} 3.9 \cdot 10^{-5} & 1.32 \cdot 10^{-6} & 2.7 \cdot 10^{-5} & 7.0 \cdot 10^{-3} \\ 2.9 \cdot 10^{-5} & 2.7 \cdot 10^{-5} & 1.3 \cdot 10^{-6} & 1.5 \cdot 10^{-3} \\ 2.0 \cdot 10^{-6} & 3.9 \cdot 10^{-5} & 2.9 \cdot 10^{-5} & 2.2 \cdot 10^{-3} \\ 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix} \quad (6)$$

The standard error $\sigma_{\hat{T}_L^C}$ in Eq. (6) quantifies the statistical uncertainty, i.e. the precision or generalization of the solution over all optimizer inputs. Each element in matrix $\hat{T}_L^C$ represents an average of that element over N cycles or time steps of Eq. (1). Similarly, we compute the variances σ^2 for each element, and consequently obtain the standard errors $\sigma_{\hat{T}_L^C} \simeq \sigma/\sqrt{N}$.

Examples of the projected LiDAR pointcloud onto the corresponding thermal image are displayed in Fig. 6.

To quantitatively evaluate the estimated extrinsic transformation T_L^C from Eq. (5), Fig. 7 presents the translational and rotational error between the approximated pose of the LiDAR in world frame $\hat{T}_L^{(i)} = \left((T_L^C)^{-1} \cdot (T_C^{(i)})^{-1} \right)^{-1}$ and the observed pose $T_L^{(i)}$. The dashed red line in both graphs represents the mean norm of the error in translation and the mean angle error in rotation.

Limitations. In Fig. 7, the errors visible are not constant over time. These follow not only from the pose estimation errors, but also from the time synchronization error when 10 Hz LiDAR scans are matched with the closest images coming from 30 Hz frequency. This error is only present when the platform is moving. Hence, it does not affect our calibration method. However, if the platform is used to capture data from motion, the time synchronization should be done properly. This method can thus also be used to check the time synchronization of the system. Furthermore, obtaining input data to the optimizer with a variety of different vertical positions is challenging due to the 16 scanning lines only covering half of the thermal image. Despite the best efforts to automatically remove outliers, some still remain as can be seen by the large error peaks in Fig. 7.

5. Conclusion

We have presented a low-cost extrinsic calibration method for thermal-LiDAR systems that employs a single calibration target: an inverted half-box with one face equipped with heating elements to produce a distinct and constant calibration pattern using thermal gradient. The three surfaces of the inverted half-box provide clear boundaries for plane extraction in the LiDAR data and to define a common world frame for both sensors. Notably, the planar surfaces of the target do not require precise orthogonality. Instead, plane normals are estimated directly from the LiDAR pointcloud, allowing for the construction of an orthonormal basis in which the thermal pattern can be aligned, for example, with the xy-plane of the coordinate system. This flexibility makes the approach particularly well-suited to low-cost implementations. Additionally, by moving the target in front of the sensor setup, the estimated extrinsic transformation is valid across the entire image frame without requiring a large target or multiple targets. These movements also enable our method to be used for verification of temporal synchronization across the thermal and LiDAR modalities. The results have been tested and validated within a range of at least five meters from the sensor setup, which is adequate for the intended indoor mapping application.

Future work will focus on enhancing the robustness of the pose estimation pipelines. For the LiDAR pose estimation, this involves improving plane segmentation, particularly in cases where the plane is only partially within the LiDAR's vertical field-of-view, as well as refining the outlier rejection process to eliminate incorrectly identified planes. In the case of the thermal camera, pose estimation can be improved by extending the effective range of the pattern detection algorithm, enabling more reliable extraction of the calibration pattern. Combined, these improvements should obtain a broader range of poses, especially in the vertical direction. Other work includes steps that have already been proposed in the literature such as implementing hardware time synchronization to reduce motion errors by

introducing a common clock, and extend the extrinsic calibration method to allow for a RGB camera and Inertial Measurement Unit (IMU) to be added into the sensor setup.

References

- Borrmann, D., Afzal, H., Elseberg, J., Nüchter, A., 2012. Mutual Calibration for 3D Thermal Mapping. *IFAC Proceedings Volumes*, 45(22), 605–610. 10th IFAC Symposium on Robot Control.
- Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., Zhang, Q., 2018. JAX: composable transformations of Python+NumPy programs.
- Dalirani, F., Heidari, F., Rahman, T., Cheema, D. S., Bauer, M. A., 2023. Automatic extrinsic calibration of thermal camera and lidar for vehicle sensor setups. *2023 IEEE Intelligent Vehicles Symposium (IV)*, 1–7.
- Fu, T., Yu, H., Yang, W., Hu, Y., Scherer, S., 2022. Targetless extrinsic calibration of stereo, thermal, and laser sensors in structured environments. *IEEE Transactions on Instrumentation and Measurement*, 71, 1–11.
- Guindel, C., Beltrán, J., Martín, D., García, F., 2017. Automatic extrinsic calibration for lidar-stereo vehicle sensor setups. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 1–6.
- Liu, D. C., Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1), 503–528.
- Mharolkar, S., Zhang, J., Peng, G., Liu, Y., Wang, D., 2022. Rgbdtcalibnet: End-to-end online extrinsic calibration between a 3d lidar, an rgb camera and a thermal camera. *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 3577–3582.
- Pusztai, Z., Hajder, L., 2017. Accurate calibration of lidar-camera systems using ordinary boxes. *Proceedings of the IEEE international conference on computer vision workshops*, 394–402.
- Valdenebro, A. G., 2016. Visualizing rotations and composition of rotations with the Rodrigues vector. *European Journal of Physics*, 37(6), 065001.
- Yang, W., Luo, H., Tse, K.-w., Hu, H., Liu, K., Li, B., Wen, C.-Y., 2024. Autonomous Targetless Extrinsic Calibration of Thermal, RGB, and LiDAR Sensors. *IEEE Transactions on Instrumentation and Measurement*.
- Yeong, D. J., Velasco-Hernandez, G., Barry, J., Walsh, J., 2021. Sensor and Sensor Fusion Technology in Autonomous Vehicles: A Review. *Sensors*, 21(6). <https://www.mdpi.com/1424-8220/21/6/2140>.
- Yuan, S., Xie, T., Zhu, S., Chen, Y., Li, Y., Zheng, T., Gu, J., 2022. A target-based calibration method for lidar-visual-thermal multi-sensor system. *2022 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 1115–1120.
- Zhang, J., Siritanawan, P., Yue, Y., Yang, C., Wen, M., Wang, D., 2018. A two-step method for extrinsic calibration between a sparse 3d lidar and a thermal camera. *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, IEEE, 1039–1044.
- Zhang, J., Zhang, R., Yue, Y., Yang, C., Wen, M., Wang, D., 2019. Slat-calib: Extrinsic calibration between a sparse 3d lidar and a limited-fov low-resolution thermal camera. *2019 IEEE international conference on robotics and biomimetics (ROBIO)*, IEEE, 648–653.