A Joint Extrinsic Calibration Tool for Radar, Camera and Lidar

Domhof, Joris; Kooij, Julian F.P.; Gavrila, Dariu M.

DOI
10.1109/TIV.2021.3065208

Publication date
2021

Document Version
Final published version

Published in
IEEE Transactions on Intelligent Vehicles

A Joint Extrinsic Calibration Tool for Radar, Camera and Lidar

Joris Domhof, Julian F. P. Kooij, and Dariu M. Gavrila

***Abstract*—We address joint extrinsic calibration of lidar, camera and radar sensors. To simplify calibration, we propose a single calibration target design for all three modalities, and implement our approach in an open-source tool with bindings to Robot Operating System (ROS). Our tool features three optimization configurations, namely using error terms for a minimal number of sensor pairs, or using terms for all sensor pairs in combination with loop closure constraints, or by adding terms for structure estimation in a probabilistic model. Apart from *relative calibration* where relative transformations between sensors are computed, our work also addresses *absolute calibration* that includes calibration with respect to the mobile robot's body. Two methods are compared to estimate the body reference frame using an external laser scanner, one based on markers and the other based on manual annotation of the laser scan. In the experiments, we evaluate the three configurations for *relative calibration*. Our results show that using terms for all sensor pairs is most robust, especially for lidar to radar, when minimum five board locations are used. For *absolute calibration* the median rotation error around the vertical axis reduces from 1° before calibration, to 0.33° using the markers and 0.02° with manual annotations.**

***Index Terms*—Calibration, camera, intelligent vehicles, lidar, optimization, ROS, radar, robots.**

I. Introduction

NOWADAYS, mobile robots have sensor setups consisting of multiple sensors for environmental perception. To increase robustness, these sensor setups consist of various sensing modalities such as lidars, cameras and radars [1], [2]. For effective sensor data fusion, a geometrical description is needed that describes the location and orientation of all the robot's sensors with respect to each other, and to its body. For that, all sensors need to be calibrated.

One can distinguish two types of calibration tasks, namely intrinsic calibration and extrinsic calibration. Intrinsic calibration involves estimating the internal parameters of the sensor. For a camera, this calibration procedure consists of estimating all entries of the camera projection matrix (focal length, skew parameter and principal point [3]) and the distortion coefficients of the lens. For a lidar, the intrinsic parameters are range offset, scale factor, vertical offset, elevation angle and azimuth angle [4]. Extrinsic calibration instead estimates the orientation and the position of the sensor (i.e. sensor pose) with respect to a frame of reference, which is also called pose estimation [5] and sensor registration [6].

Extrinsic calibration methods can further be split into two groups: target-less and target-based methods. Target-less methods (e.g. [7]–[9]) are potentially able to perform online calibration as these methods use natural features in the environment to calibrate the sensors. However, target-less calibration methods are challenging since these methods need to deal with asynchronous and heterogeneous sensors. Target-based methods instead use specifically designed physical calibration objects (i.e. targets) to obtain robust features. A typical example of a calibration target is the checkerboard pattern for intrinsic and extrinsic (stereo) calibration of cameras [3], [5], [10]. Since each sensing modality (lidar, camera and radar) works on a different wavelength and operating principles, it is challenging to find corresponding features across sensing modalities. Therefore, we focus on target-based procedures to obtain accurate key points for all involved sensors at once. Multiple correspondences can be found by repositioning the calibration target at various locations in the overlapping Field of View (FOV) of the sensors.

While reasonable initial estimates of all sensor poses can be obtained from technical drawings of the robot (e.g. computer-aided design (CAD) models), an extrinsic calibration considers the sensor measurements to determine their actual poses. In this work, we consider a rigid robot body, which means that the transformations between the sensors and the body coordinate frame are constant (i.e. no relative movement). Extrinsic sensor calibration can be split into two procedures: First, a *relative calibration* procedure estimates the sensor poses relative to all other sensors, see Fig. 1. Second, an *absolute calibration* procedure estimates sensor poses with respect to a body coordinate frame of the robot. If a *relative calibration* is done first, the *Absolute calibration* only needs to estimate the transformation of one sensor to the robot body to complete the geometric model.

Existing multi-modal calibration methods usually only address combinations of two sensor sensing modalities. Accordingly, each approach uses a calibration target design that only works for their sensor pair, e.g. lidar and monocular camera. For more complex sensors setups involving radar, camera and lidar calibration, such as intelligent vehicles, multiple calibration boards and calibration tools would be needed to calibrate all

Manuscript received February 21, 2020; accepted May 28, 2020. Date of publication March 17, 2021; date of current version August 23, 2021. This work was supported by NWO TTW under the project STW#13434 Standardized Self-Diagnostic Sensing Systems for Highly Automated Driving. *(Corresponding author: Joris Domhof.)*

The authors are with the Intelligent Vehicles group, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: jfmdomhof@protonmail.com; j.f.p.kooij@tudelft.nl; d.m.gavrila@tudelft.nl).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIV.2021.3065208.

Digital Object Identifier 10.1109/TIV.2021.3065208

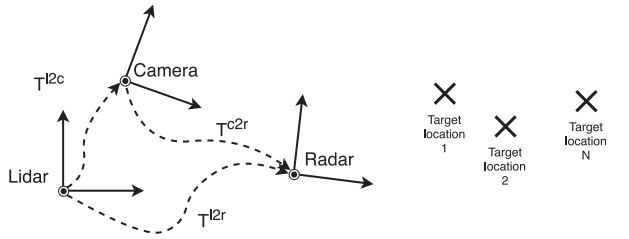


Fig. 1. Schematic overview of an example sensor setup with three coordinate frames (lidar, camera and radar) with transformation matrices from one reference frame to another, e.g. $l2c$ for lidar to camera. Joint multi-sensor calibration requires detections from multiple target locations which can be detected by all sensors simultaneously.

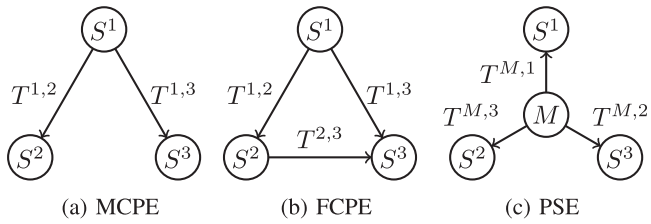


Fig. 2. Optimization configurations for joint calibration. The symbols S^i stand for sensor reference frames, and $T^{i,j}$ for coordinate frame transformations from sensor reference frame i to j. 2(a) Minimally connected pose estimation (MCPE) relies on a reference sensor S^1; 2(b) Fully connected pose estimation (FCPE) adds the loop constraint $T^{2,3} \cdot T^{1,2} = T^{1,3}$. 2(c) Pose and structure estimation (PSE) also estimates latent variables M that represent the true board locations (i.e. the structure).

sensors. However, as we will show, optimization of all sensor pairs jointly should be preferred over separate pairwise calibration. Furthermore, a joint extrinsic calibration procedure reduces the calibration effort and calibration time, since the sensors poses are estimated at once using a single calibration target design. Related work also typically only addresses *relative calibration*, while in practice *absolute calibration* is often needed.

In this work, we focus on a joint extrinsic sensor calibration procedure for sensor setups containing lidars, radars and/or cameras, using a single target design for all these sensing modalities. We consider three configurations to jointly calibrate such multi-modal setups, as shown in Fig. 2: *Minimally Connected Pose Estimation* (MCPE) estimates sensor-to-sensor transformations with respect to a single reference sensor. *Fully Connected Pose Estimation* (FCPE) provides transformations between all sensor pairs by adding a constraint that forces loop closure. The configuration *Pose and Structure Estimation* (PSE) jointly estimates sensor poses as well as the structure (i.e. calibration board poses). Additionally, we address the problem of target-based *absolute calibration* to relate the sensors to a robot's body coordinate frame. Our work is implemented in an open-source tool with bindings to Robot Operating System (ROS).

The next section addresses the related work in detail. After that, our proposed approach is presented that elaborates on the three joint calibration configurations and the procedure to calibrate the sensors with respect to the robot coordinate frame. Finally, the experimental section provides comparisons of these three configurations on real sensor data from a sensor setup with a lidar, a stereo camera and a radar. Furthermore, the two methods are evaluated to determine the body reference frame for *absolute calibration*.

II. Related Work

An overview of related work on multi-modal extrinsic calibration is provided in Table I, which is elaborated on in the following subsections. Note that a sensor pair with a stereo camera could be calibrated as two separate monocular cameras, however this is suboptimal if a full point cloud of the stereo camera is available (i.e. in case of a calibrated stereo camera).

A. Pairwise Calibration

The method of Peršić *et al.* [11] focuses on lidar to radar calibration. Rectangular shaped objects are inaccurate to detect in a lidar sensor, because nearly vertical or horizontal edges might fall between lidar scan planes (finite resolution issues). Therefore the authors use a triangular shaped Styrofoam calibration target with an attached metal trihedral corner reflector. Corner reflectors are a common target for radar because of their distinct reflectivity, the Radar Cross Section (RCS) value. The reprojection error between point cloud data and radar detections is minimized in their optimization procedure. In addition, the RCS values of multiple target locations are used to refine a subset of the transformation parameters.

Lidar to stereo calibration can be performed using the method of Guindel *et al.* [12]. This method uses a calibration target with four circles to calibrate a lidar and a stereo camera. Iterative Closest Point (ICP) [28] minimizes the error between the detected circle centers in both sensors.

For lidar to monocular camera calibration there are more methods available, namely [4], [8], [13]–[20]. Mirzaei *et al.* [4] perform intrinsic calibration of the lidar as well extrinsic calibration with respect to a monocular camera. The authors refine an analytical solution for intrinsic and extrinsic parameters by an optimization procedure based on iterative least squares. Geiger *et al.* [14] use data from multiple checkerboard patterns that are positioned in the environment to calibrate a lidar and a monocular camera. A set of initial transformation hypothesis are generated by a global registration procedure that minimizes the distance between the normal vectors and the centroids of the checkerboard patterns. After that, the set of transformation hypothesis is refined using ICP that minimizes the sum of point-to-point distances.

Extrinsic calibration of radar and monocular camera is performed by several methods [11], [21]–[23], [29]. El Natour *et al.* [21] solve a system of equations with additional spherical and geometrical constraints to obtain the transformation matrix. Both [22] and [23] estimate a homography projection between the two sensors, which means that the full 3D transformation is not available.

TABLE I

RELATED WORK ON MULTI-MODAL EXTRINSIC SENSOR CALIBRATION. THE ABBREVIATIONS IN COLUMN *OPTIMIZATION (OPTIM.)* DENOTES THE OPTIMIZATION PROCEDURE WHERE *PAIRWISE* REFERS TO OPTIMIZATION OF THE TRANSFORMATION BETWEEN A PAIR OF SENSORS AND WHERE *JOINT* REFERS TO JOINT OPTIMIZATION OF THE WHOLE POSE GRAPH. IN ADDITION, THE COLUMN *ABS./REL.* INDICATES IF THE WORK CONSIDERS *ABSOLUTE CALIBRATION* OR *RELATIVE CALIBRATION*. FURTHERMORE, THE LETTERS *L, R, S* AND *M* STAND FOR LIDAR, RADAR, STEREO CAMERA AND MONOCULAR CAMERA, RESPECTIVELY. FOR INSTANCE, THE COLUMN L & M STANDS FOR CALIBRATION OF THE SENSOR PAIR OF LIDAR AND MONOCULAR CAMERA. SYMBOLS ✓ AND ✗ INDICATE WHETHER THE EXPERIMENTS, DOCUMENTATION OR SOFTWARE SHOW THAT THE WORK CAN CALIBRATE A PARTICULAR SENSOR PAIR. SYMBOL ∼ INDICATES THAT A SENSOR PAIR WITH A STEREO CAMERA COULD BE CALIBRATED AS TWO SEPARATE MONOCULAR CAMERAS, IN PRINCIPLE. THE COLUMN *SW* INDICATES IF THE SOFTWARE IS OPEN-SOURCE AND AVAILABLE TO THE COMMUNITY.

	Configuration	Optim.	Abs./Rel.	L & R	L & S	L & M	R & S	R & M	SW	Toolbox name
Peršić et al. [11]	MCPE	Pairwise	Rel.	✓	✗	✗	∼	✓	✗	
Guindel et al. [12]	MCPE	Pairwise	Rel.	✗	✓	✗	✗	✗	✓	velo2cam_calibration
Chen and Chien [13]	MCPE	Pairwise	Rel.	✗	✓	✓	✗	✗	✗	
Geiger et al. [14]	MCPE	Pairwise	Rel.	✗	∼	✓	✗	✗	∼	Online web toolbox
Velas et al. [15]	MCPE	Pairwise	Rel.	✗	∼	✓	✗	✗	✓	but_calibration_camera_velodyne
Alismail et al. [16]	MCPE	Pairwise	Rel.	✗	∼	✓	✗	✗	✓	calidar (MATLAB)
Zhang & Pless [17]	MCPE	Pairwise	Rel.	✗	∼	✓	✗	✗	✓	RADLOCC (MATLAB)
Dhall et al. [18]	MCPE	Pairwise	Rel.	✗	∼	✓	✗	✗	✓	lidar_camera_calibration
Mirzaei et al. [4]	MCPE	Pairwise	Rel.	✗	∼	✓	✗	✗	✗	
Gong et al. [19]	MCPE	Pairwise	Rel.	✗	∼	✓	✗	✗	✗	
Vasconcelos et al. [20]	MCPE	Pairwise	Rel.	✗	∼	✓	✗	✗	✗	
El Natour et al. [21]	MCPE	Pairwise	Rel.	✗	✗	✗	∼	✓	✗	
Sugimoto et al. [22]	MCPE	Pairwise	Rel.	✗	✗	✗	✗	✓	✗	
Wang et al. [23]	MCPE	Pairwise	Rel.	✗	✗	✗	✗	✓	✗	
Apollo [24]	MCPE	Pairwise	Rel.	✗	∼	✓	∼	✓	∼	Only executables
Sim et al. [25]	MCPE/FCPE	Joint	Rel.	✗	∼	✓	✗	✗	✗	
Pusztai et al. [26]	PSE	Joint	Rel.	✗	∼	✓	✗	✗	✗	
Owens et al. [27]	PSE	Joint	Rel.	✗	∼	✓	✗	✗	✗	
Proposed	MCPE/FCPE/PSE	Joint	Rel.&Abs.	✓	✓	✓[1]	✓	✓[1]	✓	multi_sensor_calibration

[1]Our repository contains a calibration board detector for monocular cameras, therefore sensor pairs with a monocular camera can also be calibrated.

B. Joint Calibration

In order to calibrate a multi-modal sensor setup, one could simply pairwise calibrate all sensors with respect to one reference sensor, i.e. *minimally connected pose estimation*.

Alternatively, one could get inspiration from Simultaneous Localization and Mapping (SLAM), where loop closure is applied to readjust a trajectory of poses when the robot revisits the same location [30]. *Fully connected pose estimation*, a loop closure can be added as a constraint in the optimization procedure in extrinsic sensor calibration. In case of loop closure, moving over the edges in the loop (see Fig. 2(b)) should result in the original pose, i.e. the multiplication of the transformation matrices of sensors in a loop results in the identity matrix. Sim *et al.* [25] use this 'loop closure' constraint for calibration of a lidar with multiple cameras.

Visual Odometry estimates the ego-motion based on matched features in consecutive images, and it could include bundle adjustment that refines all poses in a (sub)trajectory [31]. Bundle adjustment simultaneously refines sensor poses and 3D coordinates of landmarks [31]. A similar approach can be applied to extrinsic calibration. Pusztai *et al.* [26] uses a 'bundle adjustment-like' approach that consists of two steps, where in the first step the lidar errors are minimized and in the second step the camera re-projection errors are minimized. Owens *et al.* [27] use a graph optimization approach to calibrate a setup consisting of multiple lidars and cameras.

C. Contributions

The overview in Table I reveals several open issues: Existing work only addresses *relative calibration*, is not able to calibrate all combinations of radar, lidar, and (stereo) camera jointly, and the community lacks an open-source tool to jointly calibrate such a multi-modal sensor setup.

Our work addresses these issues with four contributions. First, we examine three extrinsic calibration configurations to jointly calibrate a sensor setup consisting of lidars, cameras and radars. Important factors like configuration choice, required number of calibration board locations and choice for the reference sensor are investigated using a real multi-modal sensor setup. Second, we propose and compare two methods to estimate the pose of the body reference frame of the robot in order to perform *absolute calibration*. Third, a calibration target design that is detectable by lidar, camera and radar is presented. Fourth, the software is released as an open-source extrinsic calibration tool with bindings to Robot Operating System (ROS)[1]. For ROS users, we also provide a tool that updates the Unified Robot Description Format (URDF) file that describes the robot model, to facilitate user-friendly usage of our tool on real robotic platforms.

This article extends our conference contribution [32], which only considered *relative calibration*. In the experiments, we have increased the number of combinations of calibration board locations. In addition to relative calibration, we now also discuss *absolute calibration*, and compare two approaches to estimate the body's reference frame using an external laser scanner. Additional outdoor experiments with a moving vehicle are performed to assess the impact of calibration.

III. PROPOSED APPROACH

In this section we present our joint extrinsic calibration tool to calibrate lidar, camera and radar jointly with respect to the body reference frame of the robot. Fig. 3 shows the pipeline with all steps to calibrate the sensors with respect to the body reference frame of the robot.

The next section discusses the calibration board design. Then, the detectors are described that extract the key points from this

[1]github.com/tudelft-iv/multi_sensor_calibration

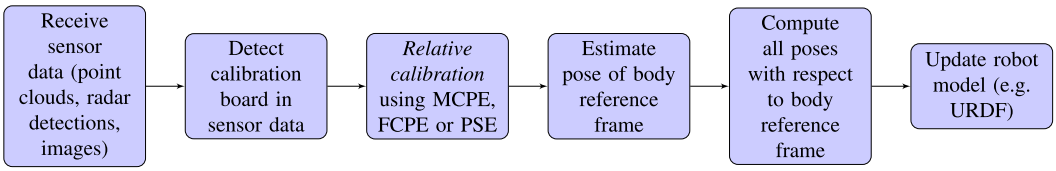


Fig. 3. Extrinsic multi-sensor calibration pipeline. The first three steps perform *relative calibration* estimating the transformation matrices between all sensors using one of three optimization configurations (MCPE, FCPE, PSE). For *absolute calibration*, the next two steps relate the sensor frames to the robot body frame, by scanning the calibration targets and the robot body with an external laser scanner. The final step updates the URDF file with the new calibration results.

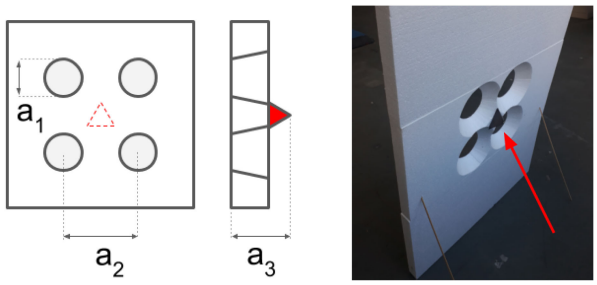


Fig. 4. From left to right, front view drawing, side view drawing, and an image of the back of the target. The trihedral corner reflector is indicated in red (triangle and arrow).

calibration board design. Using the detections, we present the details on pairwise calibration and then we extend that to joint calibration of a multi-modal sensor setup that consists of more than two sensors. The last part contains the proposed approach for *absolute calibration*.

A. Calibration Target Design

The design of the calibration target should facilitate accurate detections for all sensing modalities. For accurate radar detections, we use a trihedral corner reflector that facilitates radar reflections with specific RCS values. To limit the effect on detectability of the corner reflector, Styrofoam is chosen as material for the calibration target [33]. As target for lidar and camera, we pursue the approach of [12], [15] and use circular holes. These holes have edges, which are perfect features to detect in both sensors. The layout of the target, with a size of 1.0 m by 1.5 m, with circle diameter $a_1 = 15$ cm, and distance between the centers $a_2 = 24$ cm is shown in Fig. 4. The reflector is positioned in the middle of the four circles at the back of the Styrofoam plate (at $a_3 = 10.5$ cm from the front).[2]

B. Detection of Calibration Target

We have adapted the lidar detector and the stereo detector of [12]. For lidar and camera, the 3D location of the circle centers are returned as features. Incorrect detections can be discarded since the geometry of the board is known and there are four feature points. If the ratio between the diagonal and the side of the square is not equal to $\sqrt{2}$, detections can be discarded.

[2]See README file in the repository for details on the calibration board.

The radar measurements consist of 2D locations in polar coordinates and a RCS value. First, all detections are kept that are within the expected RCS range. From all those detections, the closest measurement to the robot is taken as radar detection as we assume that the calibration board is the closest target in the vicinity of the robot.

For the monocular camera detector, the four circles are detected based on edges in the 2D image plane. Using the known geometry of the calibration board, perspective-n-point algorithm (PnP) [34] can be used to extract the 3D locations of the circle centers.

C. Pairwise Calibration

First, we will explain pairwise calibration, which we then extend to joint calibration of a setup with N sensors in Section III-D.

The calibration target is positioned at K different locations in FOV of two sensors, referred to as sensor 1 and sensor 2. Each detector returns K detections $\boldsymbol{y}^1 = \{\boldsymbol{y}_1^1, \ldots, \boldsymbol{y}_K^1\}$ for sensor 1 and $\boldsymbol{y}^2 = \{\boldsymbol{y}_1^2, \ldots, \boldsymbol{y}_K^2\}$ for sensor 2. Each calibration board location provides four detections in 3D for lidar and camera: $\boldsymbol{y}_k = (y_{k(1)}, \ldots, y_{k(4)})$. Furthermore, the radar detector only returns a single detection as the target has one trihedral corner reflector. This detection $\boldsymbol{y}_k = (y_{k(1)})$ is defined in 2D Euclidean coordinates. Since a detector might not always detect the target, for instance if the target is not in the sensor's FOV, we use an indicator variables μ_k^i to represent if the detector of sensor i was able to successfully detect calibration board location k. This means that $\mu_k^i = 1$ if the target was detected and $\mu_k^i = 0$ otherwise.

Extrinsic calibration between the two sensors aims to estimate the relative rigid transformation $T^{1,2}$. This transformation can be used to project a point from the coordinate frame of sensor 1 to the coordinate frame of sensor 2. The rigid transformation is expressed as a 4×4 matrix for homogeneous coordinates that consists of a 3×3 rotation matrix R and 3D translation t vector,

$$T^{1,2} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}. \tag{1}$$

To use this homogeneous representation, each 3D point (x, y, z) is represented as an augmented 4D vector $(x, y, z, 1)$. To parametrize the 6 degrees of freedom of transformation $T^{1,2}$, we

use the vector $\theta^{1,2} = (t_x, t_y, t_z, v_x \cdot \alpha, v_y \cdot \alpha, v_z \cdot \alpha)$. The rotation part is expressed by an axis-angle representation (using Rodrigues' rotation formula), namely as a unit vector (v_x, v_y, v_z) for the axis of rotation, and an angle α.

For the k-th target location, the total squared Euclidean distance of the four detected circle centers is used to define the transformation error between lidar and camera detections,

$$\epsilon_k\left(\theta^{1,2}\right) = \sum_{p=1}^{4} \left\| y^2_{k(p)} - T^{1,2} \cdot y^1_{k(p)} \right\|^2. \tag{2}$$

If the sensor pair contains a radar, a different error term is used. Let $\boldsymbol{y}^R_k$ represents the radar measurement of target k, then the squared Euclidean error equals

$$\epsilon_k\left(\theta^{1,R}\right) = \left\| y^R_{k(1)} - p(T^{1,R} \cdot g(\boldsymbol{y}^1_k)) \right\|^2. \tag{3}$$

Here, function $g(\boldsymbol{y}_k)$ calculates the expected 3D position of the trihedral corner reflector in the reference frame of sensor 1 by using the four circle center locations in detection $\boldsymbol{y}_k$ and the geometry of the calibration board. Then, function $p(q_k)$ first converts 3D Euclidean point q_k to spherical coordinates (r_k, ϕ_k, ψ_k), disregards the elevation angle ψ_k, and converts (r_k, ϕ_k) back to 2D Euclidean coordinates.

In addition, we add constraints that enforce that the projected 3D points lie within radar Field of View (FOV). To achieve that, the elevation angles ψ_k for all calibration board locations k should be within the maximum view angle ψ_{max} of the radar,

$$|\psi_k| - \psi_{max} \leq 0, \quad \forall k. \tag{4}$$

Pairwise calibration is now formulated as an optimization problem that finds the optimal transformation between both sensors by minimizing the total error $f(\theta^{1,2})$ between all K calibration targets,

$$f(\theta^{1,2}) = \sum_{k=1}^{K} \mu^2_k \cdot \mu^1_k \cdot \epsilon_k\left(\theta^{1,2}\right). \tag{5}$$

The indicator variables $\mu^2_k \cdot \mu^1_k$ ensure calibration board locations K that are detected by both sensors are included. By minimizing the error criterion $f(\theta)$ subject to zero or more (in)equality constraints (e.g. equation (4)), the optimal relative transformation are obtained.

Sequential Least SQuares Programming (SLSQP) from the *SciPy* library [35] is used to solve the optimization problem, which is potentially subject to constraints. To initialize the optimization procedure, an initial solution is required. For that, the optimal rotation between the point cloud containing centroids of the four circle detections for all calibration board locations K is computed by Kabsch algorithm [36]. Using this rotation matrix, the initial translation vector can be determined. To find an initial transformation for a sensor pair containing a radar, it is assumed that detections lie on the radar plane (zero elevation angle).

D. Joint Calibration With More Than Two Sensors

To generalize extrinsic calibration from pairwise calibration to N sensors, three configurations are considered to jointly calibrate a multi-sensor sensor setup, namely MCPE, FCPE, PSE. Instead of estimating a single edge (i.e. sensor-to-sensor transformation), now multiple edges are present. The three configurations for *relative calibration* are visualized in Fig. 1 and will be discussed in this section.

a) *Minimally connected pose estimation (MCPE)*. In the first configuration, all sensors are calibrated in a pairwise manner with respect to a selected 'reference' sensor. This results in a *minimally connected* graph, which is visualized in Fig. 2(a). The edges describe the transformation from the 'reference sensor' to the other sensors. Without loss of generality, let's assume that the first sensor is selected as the reference sensor. In this case, the optimization criterion is formulated as

$$f(\boldsymbol{\theta}) = \sum_{i=2}^{N} \left[\sum_{k=1}^{K} \mu^i_k \cdot \mu^1_k \cdot \epsilon_k\left(\theta^{1,i}\right) \right]. \tag{6}$$

Note that transformations between any non-reference sensors i, j can be computed from the known transformations in this graph, i.e. $T^{i,j} = T^{1,j} \cdot (T^{1,i})^{-1}$.

b) *Fully connected pose estimation (FCPE)*. In the second configuration, we consider optimizing transformations between all sensors at once, without assigning a specific reference sensor. This results in optimizing edges in a fully connected graph (see Fig. 2(b)), akin to a loop closure optimization in SLAM. Instead of estimating $N-1$ transformation matrices with respect to a reference sensor, all transformation matrices between all $\binom{N}{2}$ combinations of two sensors are computed. In this case, the error functions equals

$$f(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left[\sum_{k=1}^{K} \mu^i_k \cdot \mu^j_k \cdot \epsilon_k\left(\theta^{i,j}\right) \right]. \tag{7}$$

To ensure that all loops l equal the identity matrix, the loop closure constraint is included in the optimization problem,

$$(T^{s_l,1} \cdot T^{s_l-1,s_l} \cdot \ldots \cdot T^{1,2}) - I = 0, \qquad \forall l \tag{8}$$

where s_l equals the number of sensors in this loop l. In this work, we only consider all $\binom{N}{3}$ combinations of $s_l = 3$ sensors. The advantage of this optimization is that it is potentially more robust against noisy observations from one reference sensor. The disadvantages are that the number of error terms increases with the number of sensors N and that by adding extra sensors, additional loop constraints must be included as well.

c) *Pose and structure estimation (PSE)*. The third configuration is called pose and structure estimation and it is visualized in Fig. 2(c). This configuration has similarities to bundle adjustment since it simultaneously estimates all sensor poses and calibration board poses. This means that both the unknown structure $M = (m_1, \ldots, m_K)$ of the true target poses in a fixed coordinate frame, and the transformation $T^{M,i}$ from the fixed frame to each sensor i are estimated. Observations are considered samples from a probabilistic measurement model, which uses $\widehat{y}^M_{k(p)} = h(m_k, p)$, with zero-mean Gaussian noise,

$$y^i_{k(p)} = T^{M,i} \cdot \widehat{y}^M_{k(p)} + \eta^i, \quad \eta^i \sim \mathcal{N}(0, \Sigma^i). \tag{9}$$

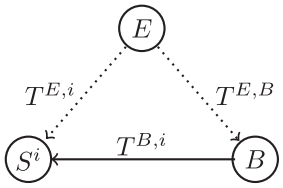


Fig. 5. For *absolute calibration*, the transformation between body reference frame B and a sensor S^i (solid arrow) is found indirectly by first determining the transformations to both frames from an external reference sensor E (dashed arrows).

Therefore, instead of the squared Euclidean distance, we use the squared Mahalanobis distance, which equals

$$D^2_{\Sigma}(a,b) = [a-b]^{\top} (\Sigma)^{-1} [a-b] \tag{10}$$

with vectors a and b, and covariance Σ. In the optimization, we jointly optimize the transformations and structure,

$$\epsilon_k \left(\theta^{M,i}, M\right) = \sum_{p=1}^{4} D^2_{\Sigma^i} \left(y^i_{k(p)}, T^{M,i} \cdot \widehat{y}^M_{k(p)}\right), \tag{11}$$

$$f(\boldsymbol{\theta}, M) = \sum_{i=1}^{N} \left[\sum_{k=1}^{K} \mu^i_k \cdot \epsilon_k \left(\theta^{M,i}, M\right)\right], \tag{12}$$

and initialize all Σ^i as identity. We use an iterative procedure to calculate the diagonal elements of the noise covariances. Using the result of the first optimization, the noise covariances are recalculated and updated, after which the optimization of $f(\boldsymbol{\theta}, M)$ is repeated. This process is continued until all variances have converged. Note that to determine a unique solution, one transformation $T^{M,i}$ must be fixed.

This probabilistic formulation has the potential advantage that it avoids having heterogeneous error functions (pixel versus Euclidean). Instead, a homogeneous error function is used that comprises of the sum of squared Mahalanobis distances. Furthermore, it provides the option to include prior knowledge on board and sensor poses, however we have not pursued this direction here. The disadvantages are also twofold. First, the optimization is more complex and therefore it takes more time. Second, the loop closure constraint is not explicitly enforced.

E. Pose Estimation of Body Reference Frame

To estimate the pose of the body reference frame of the robot, minimal three 3D reference points on the exterior of the robot are required. To determine the set of 3D points during calibration, an external sensor must be used which can detect these reference points, and the calibration target at multiple locations. This sensor should have a high resolution and large field of view to accurately locate both the 3D reference points on the exterior as well as the calibration target. From the shared detected calibration targets, the transformation from the external sensor to the robot sensor can be found, similar to for *relative calibration*. After the robot reference frame is determined in the external sensor frame too, the sought transformations between the sensor and the robot frame can be computed directly, as illustrated in Fig. 5.

To localize the 3D body reference points within the external sensor point cloud, two general approaches can be taken:

1) *Manual labeling:* The locations of the set of the 3D reference points can be manually labeled in the sensor data. These locations can be obtained by manually labeling each individual 3D reference point in the point cloud. Alternatively, multiple points can be labeled on a visible part of the robot's exterior with a specific geometric shape (e.g circular shape). After that, a geometrical shape can be fitted on the set of labeled points.
2) *Markers:* The locations of the 3D reference points can be extracted by placing physical markers that the external sensor can easily detect. This is less laborious than labeling the data afterwards, but the accuracy depends fully on how precise the markers could be placed when the calibration procedure was performed.

In practice, we use a lidar laser scanner to construct a point cloud model of the body, and either select the 3D reference points in this point cloud, or use markers that the scanner can accurately detect.

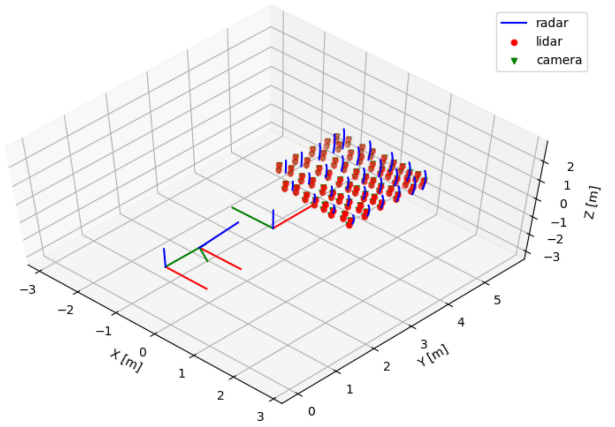
IV. Experiments

To evaluate the performance of our tool, sensor data of lidar, camera and radar is recorded with our Toyota Prius vehicle. Our Toyota Prius vehicle is equipped with:

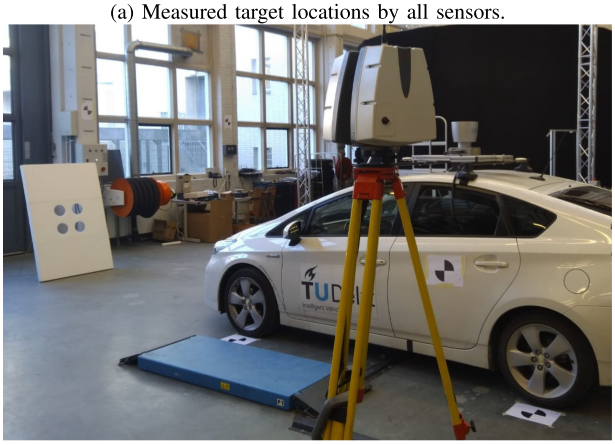
- a Velodyne HDL-64E lidar (on roof)
- a Continental ARS430 radar (behind front bumper)
- a stereo camera 2× UI-3060CP Rev. 2 (behind windscreen)

For our experimental validation, we calibrate the vehicle with the calibration target in our garage. The calibration target is positioned in front of the car at 30 different locations within approximately 5 meters. From these 30 calibration board locations, 29 locations were within the field of view of all three sensors (lidar, the stereo camera and the radar). See Fig. 6(a) for the output of our calibration tool, where the detected calibration target locations for all three sensors are shown in the lidar reference frame. For *absolute calibration*, we use a Leica P40 laser scanner as the external sensor, see Fig. 6(b). The P40 is a high resolution laser scanner which is able to localize itself in the environment using multiple black-white markers on the walls and floor. The Leica scanner was placed at several positions around the car, and using the markers and Leica software a merged point cloud of the vehicle is obtained, shown in Fig. 6(c). During calibration, the P40 is positioned next to the car such that this sensor can see both the car and 12 calibration board locations.

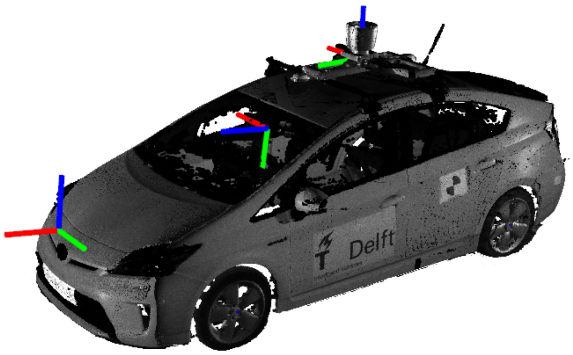
We evaluate the three configuration (MCPE, FCPE and PSE) for *relative calibration* on data from 29 calibration board locations. The computation time of the optimization depends on the number of sensors and the number of calibration board locations. If all 29 calibration board locations are used, the computation time is less than 1 s for the MCPE configuration, approximately 10 seconds for the FCPE configuration and approximately 5 minutes for PSE configuration on a high-end computer (with an Intel Xeon W-2123 @ 3.60 GHz CPU).



(a) Measured target locations by all sensors.



(b) Absolute calibration setup for a vehicle.



(c) Leica point cloud of vehicle.

Fig. 6. (a) Output of our calibration tool. All sensors poses and all detections of the calibration board are plotted in the lidar reference frame. 6(b) Absolute calibration setup, using an external Leica laser scanner, for a vehicle with a lidar, stereo camera, and radar. 6(c) Merged point cloud from the Leica scanner, with the calibrated coordinate systems of the three sensors after *absolute calibration* using the *Manual labeling* approach.

In Section IV-A, we investigate the performance of our tool for *relative calibration*, and in Section IV-B for *absolute calibration*. Finally, in Section IV-B, we present additional outdoor experiments to demonstrate the impact outside the garage in the intended environment of the vehicle.

TABLE II
COMPARISON WITH BASELINE METHOD

lidar to stereo method	# boards	RMSE [mm]
Guindel et al. [12]	single board	39.3 ± 10.4
MCPE	single board	39.3 ± 10.4
MCPE	all boards (29)	15.3
FCPE	all boards (29)	15.3
PSE	all boards (29)	15.3

TABLE III
MEDIAN OF THE RMSE [MM] FOR 200 COMBINATIONS OF 10 CALIBRATION BOARD LOCATIONS

	RMSE l2c [mm]	RMSE l2r [mm]	RMSE c2r [mm]
MCPE(camera)	16.0±0.3	20.5±0.5	27.7±0.8
MCPE(lidar)	16.0±0.3	20.4±0.5	27.6±0.7
MCPE(radar)	16.3±0.4	20.4±0.5	27.7±0.8
PSE	16.1±0.3	18.3±1.7	24.0±1.2
FCPE	16.0±0.3	15.0±0.6	22.3±0.9

A. *Relative Calibration*

To assess calibration quality, we compute for each pair of sensors the residual error, i.e. the Euclidean distance between the measured target positions after applying the found transformation to put all measurements in the same reference frame. We report the root mean squared error (RMSE) of all pairwise transformations, namely lidar to stereo camera ($l2c$), lidar to radar ($l2r$), stereo camera to radar ($c2r$). In the following sections, we compare our *relative calibration* approach to baseline calibration methods, and assess the choice of reference sensors, the number of target locations, and sensitivity to additional noise.

1) Comparison to Baseline Method: First, we compare our method with the single-target method of Guindel *et al.* [12] that only calibrates a lidar to stereo camera pair. For our MCPE implementation when using sensor data of a single target location and the single target method of Guindel, the calibration is performed for all 29 calibration board locations and the mean and standard deviation of the RMSE are provided in Table II. It can be seen that both single target implementations provide a similar result. In addition, we investigate the benefit of using multiple calibration board locations. Table II shows that the $l2c$ RMSE reduces from 39 mm to 15 mm.

2) Choice of MCPE Reference Sensor: Next, we investigate if the choice for reference sensor of the MCPE configuration influences its results. We randomly pick 200 times 10 calibration board locations and calibrate the sensors. Table III shows the median RMSE for MCPE with all three reference sensors and for FCPE and PSE. The Table shows that all choices (lidar, camera and radar) give similar RMSE, however selecting the radar as reference sensors results in two links that contain radar measurements. Since radar data is 2D (range and angle) having two links with radar data might result in less accurate results, therefore we use the lidar, with a FOV of 360°, as reference sensor from now on. Furthermore, the RMSE for the sensor pairs $l2r$ and $c2r$ shows that configurations FCPE and PSE perform better than the MCPE configuration.

3) Dependence on the Number of Calibration Board Locations: To understand the impact of the number of calibration board locations, K, we vary K from 3 to 29 locations. For

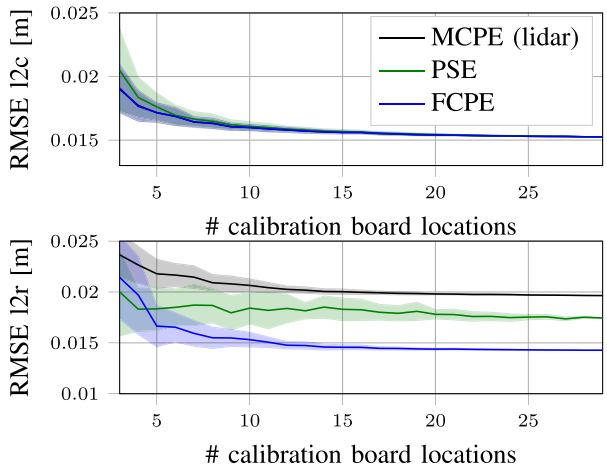


Fig. 7. The median and median absolute deviation of the RMSE on 100 board locations for varying number of calibration board locations (K).

each value of K, 100 sets of K randomly selected locations are used to calibrate the sensor setup. Fig. 7 shows the median and median absolute deviation of the RMSE over all 100 sets. Both FCPE and PSE show smaller RMSE than MCPE for the $l2r$ transform. The RMSE for $l2c$ and $l2r$ transforms for FCPE and PSE configurations have converged to ≤ 2 cm if more than 10 calibration board locations are used. The configuration FCPE shows the best performance for $l2r$, since the RMSE is smaller than 1.5 cm when using all 29 board locations.

4) Sensitivity to Observation Noise: We also compare the robustness of the three configuration under additional measurement noise for a sensor, and wonder how it affects the other sensor pairs. Zero-mean Gaussian noise $\mathcal{N}(0, \sigma^2 I_3)$ is added to the 3D measurements of the lidar detections. The median and median absolute deviation of the RMSE for various values of σ are plotted in Fig. 8, and it can be seen that the RMSE of sensor pairs with lidar increase as a result, though the $c2r$ errors for both FCPE and PSE remain fairly constant as more noise is added. Furthermore, the RMSE for $l2c$ and $l2r$ remain lower than the RMSE $c2r$ for most of values of σ.

B. Absolute Calibration

For *absolute calibration*, we seek the additional transformation between the Velodyne lidar coordinate frame and the vehicle's body reference frame using the external Leica laser scanner. This means that the transformation between the Velodyne and the body reference frame $T^{B,i}$ (see Fig. 5) needs to be assessed. In our application, the origin vehicle's body reference frame is at the center of the rear axle *projected onto the ground*, the X-axis is pointing forward, the Y-axis is pointing to the left rear wheel and the Z-axis is perpendicular to the ground (pointing upwards). Hence, to determine the pose of the body reference frame, the location of the wheel centers and ground plane must be determined.

In Section III, several practical approaches are discussed to determine 3D reference points on the body reference frame, namely *Markers* and *Manual labeling*, which we implemented as follows:

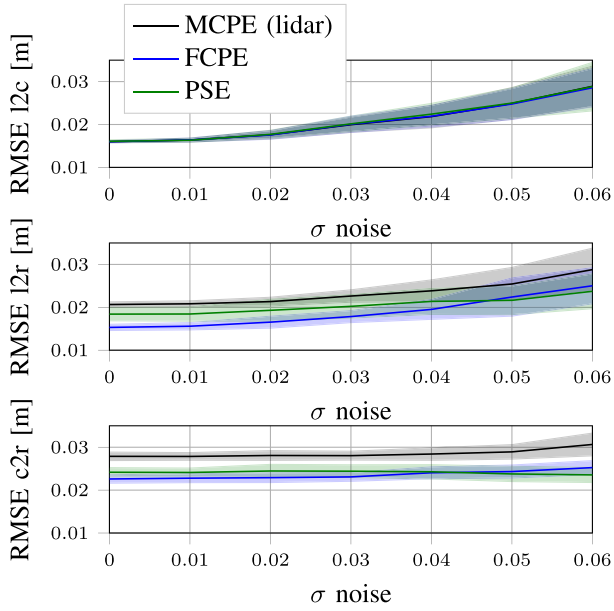


Fig. 8. RMSE error as function of Gaussian observation noise $\mathcal{N}(0, \sigma^2 I_3)$ added to the lidar observations. The plotted median and median absolute deviation are based on 100 random combinations of 10 calibration board locations.

TABLE IV
COMPARISON OF THE STANDARD DEVIATIONS OF THE WHEEL CENTER LOCATION FOR THE MANUAL LABELING APPROACHES

	σ_x [mm]	σ_y [mm]	σ_z [mm]
Single point	1.3	1.7	1.9
3D circle fit	0.8	0.5	1.3

1) As a first *Manual labeling* approach, each wheel center location is manually labeled by selecting a single point in the Leica point cloud (see Fig. 6(c)). To project those locations to the ground, the normal vector and distance to the ground is found by fitting a planar model on the lower part of the point cloud.
2) Another *Manual labeling* approach is to manually select N points on the rim of the wheel, and fit a 3D circle through those N points to determine the wheel center.
3) For the *Markers* approach, we position four Leica markers next to the wheels on the ground (see markers in Fig. 6(b)) below the axles.

This section will compare the robustness of the labeling options over multiple repetitions, and compare the rotational errors of the approaches with respect to the ground normal.

1) Robustness of Manual Labeling: First, we compare the two manual labeling approaches by labeling the left rear wheel of the car 10 times. Table IV shows the standard deviations in X, Y and Z positions in Leica reference frame. The results shows that labeling the wheel centers using multiple points ($N = 10$) on the rim and fitting a 3D circle provides slightly better results than labeling wheel centers using a single 3D point. Despite that the differences between the labeling approaches are small, we will use a 3D circle fit on the rim to determine the wheel centers from now on.

More importantly, we observe that the standard deviation between multiple annotations is in the order of millimeters. We conclude that this is sufficiently robust given the operating scale and physical size of the vehicle.

2) Rotation Error Around X-Axis and Y-Axis Combined: Now, we quantify the the error in angle between the estimated and expected Z-axis. The rotational error around the vertical Z-axis will be assessed later in Section IV-C2.

Since the Z-axis of the body reference frame is perpendicular to the ground, we expect that the observed normal vector of the ground *in the vehicle sensors* is aligned with the body's Z-axis. Our recordings were recorded in a large garage space at the same time as the absolute calibration was performed, meaning that the state of the suspension and state of the tires is unchanged. We can assume that the ground within 6 meters of the vehicle center is flat. We estimate the ground normal vector in the point cloud of the vehicle's Velodyne by segmenting the planar ground floor, using a maximum distance tolerance of 2.5 cm, and use the calibration to transform it to the body reference frame. The angular error θ between the observed normal vector n_{obs} and the expected normal vector $\mathbf{n_{exp}} = [0, 0, 1]$ is

$$\theta = \arccos\left(\frac{\mathbf{n_{obs}} \cdot \mathbf{n_{exp}}}{\|\mathbf{n_{obs}}\| \ \|\mathbf{n_{exp}}\|}\right). \tag{13}$$

Initially when the sensors were positioned based on manual adjustments, the angle θ was 0.13°, and we find that after calibration the angle θ has decreased to 0.07° using the *Markers* approach, and 0.02° using the *Manual labeling* approach.

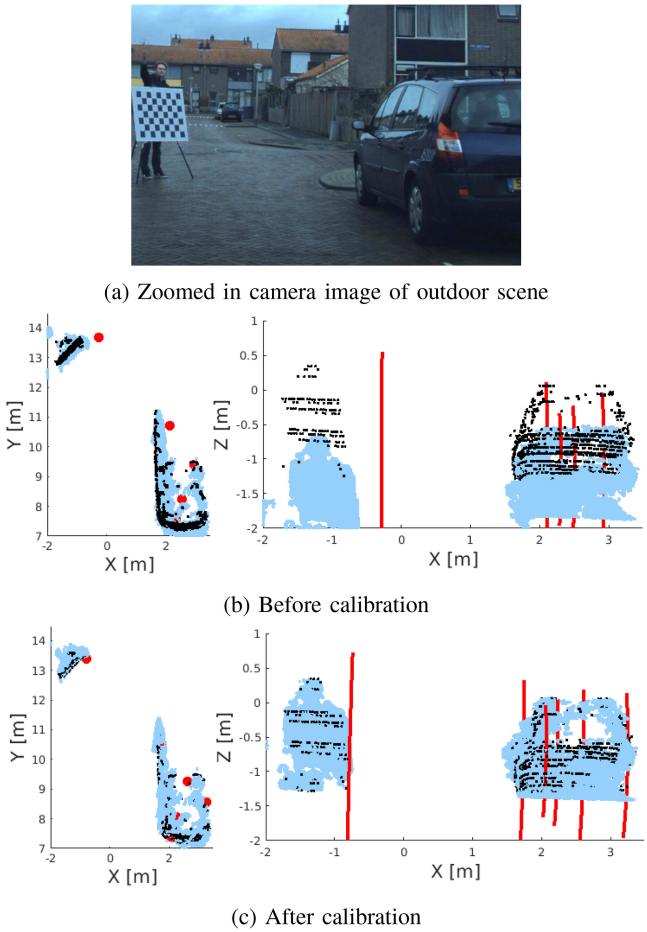
C. Outdoor Experiments

Finally, we report on additional experiments performed outside the garage at two outdoor locations. These enable us to asses the calibration impact on multi-modal perception in realistic environments, and at larger distances than possible in the garage to highlight the reduced rotational errors.

1) Location 1: Qualitative Assessment: We first qualitatively demonstrate the overall spatial and rotational accuracy of *relative calibration* for all vehicle sensors in an urban outdoor scene with obstacles at 7 to 14 m distance, see Fig. 9. Before calibration, with initial manually set sensor poses, the data from lidar and the stereo camera have a mismatch in the Z direction, and the radar detection on the person on the left does not match the measurements from the other two sensors. After calibration the data from all sensors are well aligned, even though the used calibration targets were only placed at a few meters in front of the vehicle.

2) Location 2: Rotation Error Around Vertical Z-Axis: To assess the rotation error around the vehicle's vertical axis for *absolute calibration*, we measure the apparent lateral drift in the sensor frame of static objects while the vehicle is moving straight forward, i.e. along the X-axis of the body reference frame. On a well calibrated setup, we expect that the measured lateral position of static objects, when transformed to the vehicle's body reference frame, is the same at the first and last measurement, see Fig. 10. Therefore, we measure in the vehicle's lidar the lateral positions of eight street light poles distributed along the right side of an empty 240 m long straight road. The poles are extracted from the point cloud by clustering the lidar points [37]. The car drives with a maximum speed of 5.4 m/s over the road marking line (closed road), and each pole is measured for about 30 meters. To compensate for measurement errors, small deviations of the straight trajectory, and outliers at the start and end, we fit for each pole a line through all measured positions. A pole's amount of lateral shift (ΔY) over the longitudinal range that it is observed (ΔX) allows us to compute the angular error $\alpha = \arctan(\Delta Y / \Delta X)$ of the lidar w.r.t. the body reference frame. While small deviations in the car's actual velocity can affect the number of measured positions for each street light in Fig. 11, e.g. driving faster would result in fewer measured positions for each street light, we still expect similar angle estimates as the speed only impacts the number of points that are used for line fitting.

The measured positions of the street lights in the body frame are shown in Fig. 11. We observe that the slopes for *Manual labeling* are the smallest compared to the other cases. Overall, the reported median α angles in the graph captions confirm

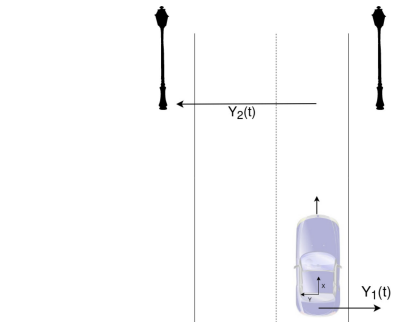


(a) Zoomed in camera image of outdoor scene

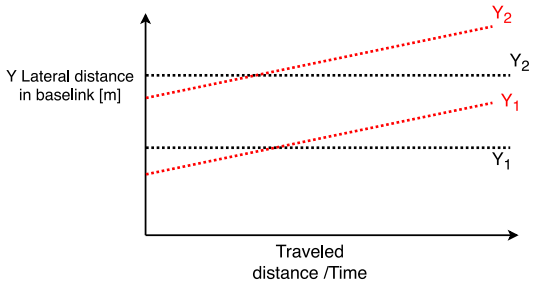
(b) Before calibration

(c) After calibration

Fig. 9. (a) Image of the recorded scene to test the calibration. There is a parked car $\sim 6\ m$ in front of the sensor setup, and a person with a checkerboard at $\sim 13\ m$. 9(b) The lidar (black) and stereo (blue) point cloud, and radar detections (red) before extrinsic calibration (based on manual adjustments). 9(c) The sensor data after extrinsic calibration. Radar detections are drawn as arcs since the elevation angle is not measured.



(a) Vehicle driving straight on road with street lights.



(b) Expected lateral position of street lights.

Fig. 10. Experimental setup to assess the rotational error of *absolute calibration* in real-world setting over larger distances. 10(a) The vehicle drives in an approximately straight line on a long straight road with streetlights. 10(b) The angle of inclination α (slope) of the lateral position of the light in the vehicle's reference frame is expected to be near zero over the whole drive if the sensors are properly calibrated (black lines). For a bad calibration (red lines) the results would show a systematic lateral drift.

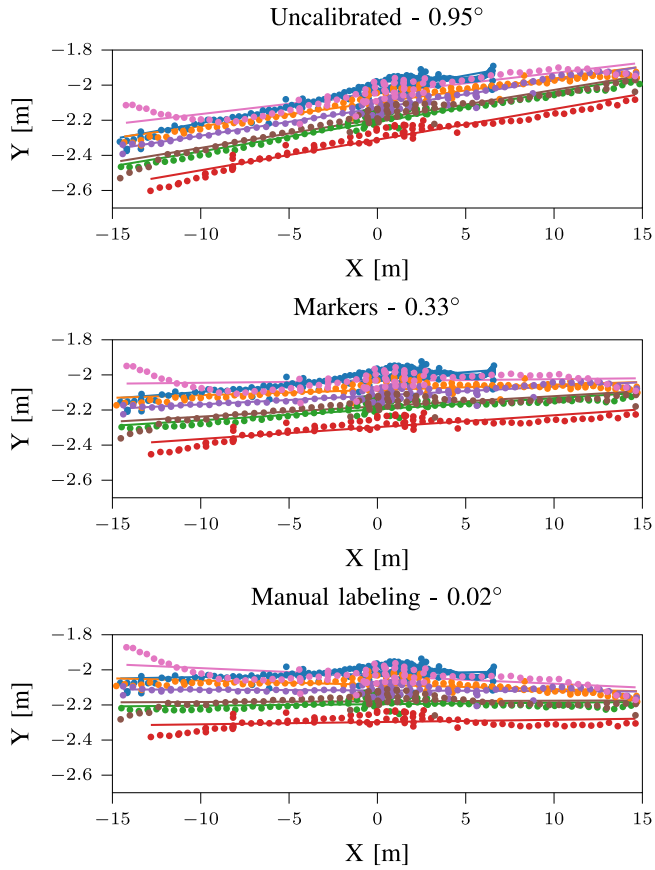


Fig. 11. Locations of all streetlights in vehicle body reference frame while driving on a straight road. The value in the title represents the median α angle for each method.

that the error has decreased from more than 0.95° to 0.33° for *Markers* and 0.02° for *Manual labeling*.

V. Discussion

Both the FCPE and the PSE configuration showed better results than the MCPE configuration (see Table III and Fig. 7). This was expected since the FCPE configuration includes all error terms between sensors in the optimization and the PSE configuration uses a probabilistic model to simultaneously estimate the calibration board poses and the sensor poses. We found that the FCPE configuration shows the best results on our sensor setup which consists of a lidar, a stereo camera and a radar. Furthermore, the experiments showed that our method that uses sensor data of multiple calibration board locations outperforms the single target method of Guindel *et al.* [12]. With more than ten calibration board locations, the median RMSE is < 2 cm for lidar to camera, approximately 2 cm for lidar to radar and approximately 2.5 cm for camera to radar. For the MCPE configuration with fast computation time, the radar does not seem to be a good choice as reference sensor, since it results in having two links with 2D radar measurements (range and angle).

The PSE configuration simultaneously estimates the calibration board poses and sensor poses. The noise covariances are estimated iteratively using sensor data of all calibration board locations, however the noise covariances might not be constant for all calibration board locations as the radar observation noise is usually larger at the edges of the field of view. It is assumed that the observations are samples from a probabilistic model with zero-mean Gaussian noise and that for every sensor the (2D/3D Euclidean) measurements are uncorrelated (i.e. all off-diagonal entries of the observation covariance matrices are equal to zero). In computation of the root mean squared error, the errors in the various dimensions (e.g. X,Y,Z) are treated equally (i.e. identity weights). These identity weights are also used in optimization of Euclidean error terms in the MCPE and the FCPE configuration. However in case of the PSE configuration, the total error term in the optimizer is based on the squared Mahalanobis distance, which means that the inverse covariance matrices are used as weights (i.e. different weights for the various dimensions). This means that in case of the PSE configuration, the total error is internally optimized using the inverse covariance matrices as weights, however when the RMSE is computed then identity weights are used. This could explain why the PSE configuration performs worse than the FCPE configuration.

Furthermore, some practical considerations are important for users. Our calibration board design consists of four key points for lidar and camera and one key point for radar (e.g. trihedral corner reflector). The number of key points for every sensor affects the optimization. In the FCPE configuration, the error term consists of all pairwise errors and the total error for a single calibration board locations consist of four error terms for lidar to camera

and one error term for the other two links. In addition, all error terms for sensor pairs with a radar are 2D Euclidean errors, whereas lidar to camera terms are 3D Euclidean errors. This means that the error in the FCPE configuration is dominated by the lidar to camera errors, since it has four 3D Euclidean errors for every calibration board location. Furthermore, there are multiple loop closure constraints for the FCPE configuration for $N > 3$ sensors. The number of constraints (loop constraints) increases with the number of sensors in the FCPE configuration. Therefore, the optimizer needs to deal with a increasing number of constraints. This might influence the performance of this configuration. In addition, the PSE configuration requires the measurement noise covariances for all sensors, therefore these are estimated in an iterative manner. In practice, this means that the computation time is significantly affected by the number of calibration board locations.

For calibration with respect to the body reference frame, we used a circle fitting approach on the rims of the wheels to determine its center, which makes this calibration approach suitable for robots with visible wheels. In absence of visible wheels, the users should use other 3D reference points to determine the pose of a body reference frame. The main difference between the approach *Markers* and the approach *Manual labeling* can be found in the rotation error around the vertical axis, which can be explained by the fact that accurate marker placement is challenging for the former method. Moreover, the accuracy completely depends on how well the markers were placed during the calibration procedure. In case of inaccurate marker placement, the calibration procedure needs be performed again. When the *Manual labeling* needs to be performed again, the point cloud model of the car (including the wheel center locations) can be reused. In that case, the transformation between the point cloud model and the current scan of the external sensor can be estimated using point set registration techniques (e.g. ICP) to determine the wheel center locations in the reference frame of the Leica. In addition, the absolute calibration of the lidar sensor was evaluated. For the lidar sensor, the method *Manual labeling* using *3D circle fitting* showed most accurate results, namely a median angle of 0.02° around the vertical Z-axis. To provide insights on how orientation errors affect position estimates at a larger distance, the displacement error due to rotation errors ϵ for objects located at distance d_{obj} can be computed using: $\Delta = \sin(\epsilon) \cdot d_{obj}$. Initially when the sensors were positioned based on manual adjustments a median angle error of 0.95° results in a displacement error of approximately 50 cm for an object at 30 meters. After calibration, the median angle reduces from approximately 1° to 0.02° with a factor 50, therefore the displacement error decreased with a factor 50 assuming small-angle approximation ($\sin(\epsilon) \approx \epsilon$ where ϵ is in radians).

VI. Conclusion

We have presented an open-source extrinsic calibration tool to jointly calibrate sensor setups consisting of lidar, camera and radar sensors. Our tool offers three configurations to estimate the sensor poses from simultaneous detections of multiple calibration board locations. Important factors like configuration choice, dependency on the number of calibration board locations and choice for the reference sensor are investigated using a real multi-modal sensor setup that consists of a lidar, a stereo camera and a radar. The experiments show that all configurations can provide good calibration results, though *fully connected pose estimation* showed the best performance. When ten calibration board locations are used, the median RMSE is less than 2 cm for lidar to camera, approximately 2 cm for lidar to radar and approximately 2.5 cm for camera to radar. Our findings highlight the importance of calibrating multiple sensors modalities jointly, rather than separately for each pair.

In addition, we described two approaches to calibrate the sensors to the body reference frame using an external laser scanner, a process referred to as *absolute calibration*. To measure the body frame pose of a vehicle in the external point cloud, we found that the best approach was to manually annotate several points on each wheel, and perform geometric shape fitting on the wheels and ground plane. For the lidar sensor, we achieved a low horizontal error w.r.t. the perceived ground plane normal, $< 0.2^\circ$, an outdoor driving experiment showed a rotation error around the vertical axis of 0.02°, an order of magnitude smaller than the alternatives.

We hope that by sharing our ROS compatible calibration tool, and detailing our approach and findings, we facilitate other researchers that need to regularly calibrate such multi-modal sensor setups.

References

[1] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1773–1795, Dec. 2013.

[2] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1239–1258, Jul. 2010.

[3] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[4] F. M. Mirzaei, D. G. Kottas, and S. I. Roumeliotis, "3D LIDAR-camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization," *Int. J. Robot. Res.*, vol. 31, no. 4, pp. 452–467, 2012.

[5] R. Szeliski, *Computer Vision: Algorithms and Applications*. Berlin, Germany: Springer Science & Business Media, 2010.

[6] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Inf. Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[7] G. Pandey, J. R. McBride, S. Savarese, and R. M. Eustice, "Automatic extrinsic calibration of vision and lidar by maximizing mutual information," *J. Field Robot.*, vol. 32, no. 5, pp. 696–722, 2015.

[8] N. Schneider, F. Piewak, C. Stiller, and U. Franke, "RegNet: Multimodal sensor registration using deep neural networks," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2017, pp. 1803–1810.

[9] J. Levinson and S. Thrun, "Automatic online calibration of cameras and lasers," in *Robotics: Science and Systems*, vol. 2, 2013.

[10] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[11] J. Peršić, I. Marković, and I. Petrović, "Extrinsic 6DoF calibration of a radar-lidar-camera system enhanced by radar cross section estimates evaluation," *Robot. Auton. Syst.*, vol. 114, pp. 217–230, 2019.

[12] C. Guindel, J. Beltrán, D. Martín, and F. García, "Automatic extrinsic calibration for lidar-stereo vehicle sensor setups," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst.*, 2017, pp. 1–6.

[13] C.-Y. Chen and H.-J. Chien, "Geometric calibration of a multi-layer LiDAR system and image sensors using plane-based implicit laser parameters for textured 3-D depth reconstruction," *J. Vis. Commun. Image Representation*, vol. 25, no. 4, pp. 659–669, 2014.

[14] A. Geiger, F. Moosmann, Ö. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2012, pp. 3936–3943.
[15] M. Velas, M. Španěl, Z. Materna, and A. Herout, "Calibration of RGB camera with velodyne LiDAR," in *Comm. Papers Proc. Int. Conf. Comput. Graph., Visual. and Comput. Vis. (WSCG)*, 2014, pp. 135–144.
[16] H. Alismail, L. D. Baker, and B. Browning, "Automatic calibration of a range sensor and camera system," in *Proc. IEEE 2nd Int. Conf. 3D Imag., Model., Process., Visual. Transmiss.*, 2012, pp. 286–292.
[17] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IEEE Cat. No. 04CH37566)*, vol. 3, 2004, pp. 2301–2306.
[18] A. Dhall, K. Chelani, V. Radhakrishnan, and K. M. Krishna, "LiDAR-camera calibration using 3D-3D point correspondences," 2017, *arXiv:1705.09785*.
[19] X. Gong, Y. Lin, and J. Liu, "3D LIDAR-camera extrinsic calibration using an arbitrary trihedron," *Sensors*, vol. 13, no. 2, pp. 1902–1918, 2013.
[20] F. Vasconcelos, J. P. Barreto, and U. Nunes, "A minimal solution for the extrinsic calibration of a camera and a laser-rangefinder," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2097–2107, Nov. 2012.
[21] G. E. Natour, O. Ait-Aider, R. Rouveure, F. Berry, and P. Faure, "Toward 3D reconstruction of outdoor scenes using an MMW radar and a monocular vision sensor," *Sensors*, vol. 15, no. 10, pp. 25 937–25967, 2015.
[22] S. Sugimoto, H. Tateda, H. Takahashi, and M. Okutomi, "Obstacle detection using millimeter-wave radar and its visualization on image sequence," in *Proc. IEEE 17th Int. Conf. Pattern Recognit.*, 2004, pp. 342–345.
[23] T. Wang, N. Zheng, J. Xin, and Z. Ma, "Integrating millimeter wave radar with a monocular vision sensor for on-road obstacle detection applications," *Sensors*, vol. 11, no. 9, pp. 8992–9008, 2011.
[24] "Apollo an open autonomous driving platform," Jul. 2019. [Online]. Available: https://github.com/ApolloAuto/apollo.
[25] S. Sim, J. Sock, and K. Kwak, "Indirect correspondence-based robust extrinsic calibration of lidar and camera," *Sensors*, vol. 16, no. 6, 2016, Art. no. 933.
[26] Z. Pusztai, I. Eichhardt, and L. Hajder, "Accurate calibration of multi-lidar-multi-camera systems," *Sensors*, vol. 18, no. 7, 2018, Art. no. 2139.
[27] J. L. Owens, P. R. Osteen, and K. Daniilidis, "MSG-cal: Multi-sensor graph-based calibration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 3660–3667.
[28] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
[29] D. Gao, J. Duan, X. Yang, and B. Zheng, "A method of spatial calibration for camera and radar," in *Proc. IEEE 8th World Congr. Intell. Control Automat.*, 2010, pp. 6211–6215.
[30] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robot. Automat. Mag.*, vol. 13, no. 2, pp. 99–110, Jun. 2006.
[31] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *IEEE Robot. Automat. Mag.*, vol. 18, no. 4, pp. 80–92, Dec. 2011.
[32] J. Domhof, J. F. Kooij, and D. M. Gavrila, "An extrinsic calibration tool for radar, camera and Lidar," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2019, pp. 8107–8113.
[33] J. Peršić, I. Marković, and I. Petrović, "Extrinsic 6DoF calibration of 3D lidar and radar," in *Proc. IEEE Eur. Conf. Mobile Robots*, 2017, pp. 1–6.
[34] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
[35] E. Jones *et al.* "SciPy: Open source scientific tools for python," 2001. Accessed: Sep. 2018. [Online]. Available: https://www.scipy.org/
[36] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Cryst.lographica Sect. A: Cryst. Phys., Diffraction, Theor. Gen. Crystallogr.*, vol. 32, no. 5, pp. 922–923, 1976.
[37] I. Bogoslavskyi and C. Stachniss, "Efficient online segmentation for sparse 3D laser scans," *PFG - J. Photogrammetry, Remote Sens. Geoinformation Science*., pp. 1–12, 2017.

Joris Domhof received the bachelor's degree in 2011 in aerospace engineering and the master's degree in 2015 in mechanical engineering from the Delft University of Technology, Delft, The Netherlands, where he is currently working toward the Ph.D. degree. His research focuses on sensor data fusion techniques for intelligent vehicles with an emphasis on detection and tracking of road users by using multiple sensing modalities which include radar, camera, and lidar.

Julian F. P. Kooij received the Ph.D. degree in 2015 in artificial intelligence from the University of Amsterdam, Amsterdam, The Netherlands, where he worked on unsupervised machine learning and predictive models of pedestrian behavior. In 2013, he was with Daimler AG on path prediction of vulnerable road users for highly-automated vehicles. In 2014, he joined the Computer Vision Lab, Delft University of Technology, Delft, The Netherlands, where since 2016, he has been an Assistant Professor with the Intelligent Vehicles Group, part of Cognitive Robotics Department. His research interests include developing novel probabilistic models and machine learning techniques to infer and anticipate critical traffic situations from multimodal sensor data.

Dariu M. Gavrila received the Ph.D. degree in computer science from the University of Maryland, College Park, MD, USA, in 1996. From 1997 until 2016, he was with Daimler R&D, Ulm, Germany, where he became a Distinguished Scientist. In 2016, he moved to Delft University of Technology, Delft, The Netherlands, where he since heads the Intelligent Vehicles Group as a Full Professor. His research deals with sensor-based detection of humans and analysis of behavior, most recently in the context of the self-driving car in complex urban traffic. He was the recipient of the Outstanding Application Award 2014 and the Outstanding Researcher Award 2019, both from the IEEE Intelligent Transportation Systems Society.