2019 International Conference on Robotics and Automation (ICRA)
Palais des congres de Montreal, Montreal, Canada, May 20-24, 2019

An Extrinsic Calibration Tool for Radar, Camera and Lidar

Joris Domhof[1], Julian F. P. Kooij[1] and Dariu M. Gavrila[1]

***Abstract*— We present a novel open-source tool for extrinsic calibration of radar, camera and lidar. Unlike currently available offerings, our tool facilitates *joint* extrinsic calibration of all three sensing modalities on multiple measurements. Furthermore, our calibration target design extends existing work to obtain simultaneous measurements for all these modalities. We study how various factors of the calibration procedure affect the outcome on real multi-modal measurements of the target. Three different configurations of the optimization criterion are considered, namely using error terms for a minimal amount of sensor pairs, or using terms for all sensor pairs with additional loop closure constraints, or by adding terms for structure estimation in a probabilistic model. The experiments further evaluate how the number of calibration boards affect calibration performance, and robustness against different levels of zero mean Gaussian noise. Our results show that all configurations achieve good results for lidar to camera errors and that *fully connected pose estimation* shows the best performance for lidar to radar errors when more than five board locations are used.**

I. Introduction

Mobile robots are often equipped with multiple heterogeneous sensors, enabling perception in various sensing modalities. For instance, automated vehicles heavily rely on lidars, radars and cameras for environmental perception to improve robustness [1], [2]. In order to represent sensor observations in a common reference frame, the rigid transformations (i.e. 3D rotation and translation) between all sensors coordinate frames must be known. A rough estimate could be obtained by assessing the sensor placement itself, e.g. with Computer-Aided Design models. However, extrinsic calibration can provide more precise estimates, as it aligns corresponding sensor measurements of real targets directly (see figure 1).

Up till now, existing calibration tools only addressed pairwise sensor calibrations of maximally two sensing modalities [3]–[12]. Since each modality has a different measurement principle, each proposed calibration procedure used different target designs. In sensor setups with more modalities, this results in duplicated calibration efforts, which is especially concerning for mobile robots at which sensors are frequently dismounted or repositioned. We instead pursue joint extrinsic calibration using a single target design for more than two modalities, namely lidar, camera, and radar. Apart from reduced effort, a single target enables us to optimize all sensors jointly which could improve accuracy and robustness. We therefore present an open-source tool and a suitable calibration target design to facilitate extrinsic calibration of three sensing modalities. The calibration tool has bindings with the commonly used Robotics Operating System (ROS).

[1] Intelligent Vehicles group, Cognitive Robotics Dept., Delft University of Technology, the Netherlands; j.f.m.domhof@tudelft.nl, j.f.p.kooij@tudelft.nl, d.m.gavrila@tudelft.nl

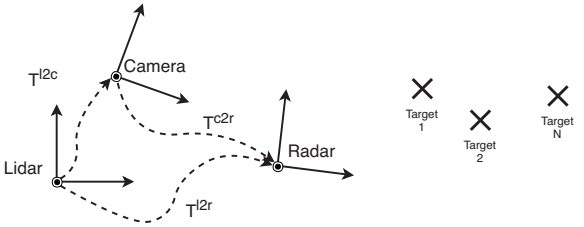
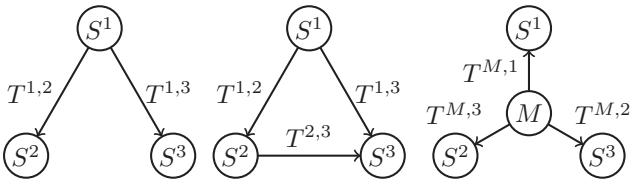


Fig. 1: Schematic overview of three reference frames: lidar, camera and radar with transformation matrices from one reference frame to another, e.g. $l2c$ for lidar to camera. Joint multi-sensor calibration requires multiple targets which can be detected by all sensors simultaneously.



(a) MCPE (b) FCPE (c) PSE

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

Since our sensor setup consists of more than two sensing modalities, we are faced with multiple options on how to formulate the optimization problem over all sensor-to-sensor transformations. Based on a review of existing methods and other related techniques (see Section II), we identify three possible configurations, as shown in figure 2: *Minimally connected pose estimation* (MCPE) estimates transformations for all sensors to a single reference sensor. *Fully connected pose estimation* (FCPE) all transformations between all sensors jointly, and enforces a loop closure constraint to ensure their consistency. *Pose and structure estimation* (PSE) instead optimizes transformations to an explicit estimate of all calibration target poses (i.e. the structure).

After discussing the related work, we shall formalize these options mathematically, and present experiments to illustrate the benefits of our tool, and investigate factors that affect the calibration result.

978-1-5386-6027-0/19/$31.00 ©2019 IEEE

II. Related Work

A first distinction in calibration methods is the difference between intrinsic and extrinsic calibration. Intrinsic calibration estimates the operational parameters of a single sensor. For instance, intrinsic camera calibration is the process of estimating focal length, the principal point, skewness and lens distortion parameters [13]. Likewise, the intrinsic lidar parameters for a single laser beam include range offset, scale factor, vertical offset, elevation angle and azimuth angle [11]. On the other hand, extrinsic calibration concerns estimating the spatial transformation of the sensor coordinates to other sensors or reference frames, and this is our main focus here. This problem is also known as pose estimation [14] and sensor registration [15].

Extrinsic calibration can in turn be divided into target-less and target-based procedures. Target-less methods [16], [17] use natural environmental features as an input to the calibration procedure. However, robustly finding corresponding natural features in all sensing modalities to jointly calibrate all sensors is challenging, since each sensor works with a different sensing principle. Target-based procedures avoid this difficulty by using a calibration target that is designed to facilitate robust and accurate feature extraction. A checkerboard pattern is a common example of a camera calibration target, as it facilitates corner detection. By moving the calibration pattern around to various locations in Field of View (FOV) of the sensors, multiple target detections can be obtained with known correspondences.

Table I shows an overview of extrinsic calibration procedures, specifically those concerning lidar, radar, and camera. It reveals that the available tools focus on calibration of two sensing modalities only. In addition, existing methods use either meters or pixels as error measures. However, robots are equipped with heterogeneous sensors, which should result in minimization of a heterogeneous error function. We will discuss the relevant methods for pair-wise cross-modal calibration now in more detail.

A. Multi-modal calibration with lidar, camera and radar

There are many methods to estimate the relative pose of a camera with respect to a lidar sensor [3], [4], [6]–[8], [10], [17]. Dhall et al. [10] use a square plate as calibration target, which has an ArUco marker (a square black-white pattern) [18] to facilitate pose estimation in monocular camera. The corners of the plate are features for both lidar and camera, and the Kabsch algorithm [19] is used to estimate the transformation between the sensors. Alismail et al. [7] propose a method to estimate both intrinsic and extrinsic parameters using point-to-plane Iterative Closest Point (ICP) [20], and Levenberg-Marquardt (LM) optimization for the rigid transformation. Geiger et al. [4] have developed a calibration method to calibrate a lidar and a camera using multiple checkerboard patterns as calibration targets. Fine registration is performed by first minimizing point-to-plane distances using the normal vectors and the centroids of the checkerboard patterns, and then minimizing point-to-point distances using ICP. Raw lidar and vision data can be uploaded to a publicly available web toolbox that performs intrinsic (camera) and extrinsic calibration. [11] estimates the intrinsic lidar parameters, and its extrinsic calibration with respect to a camera. The authors argue that the non-linear estimation problem requires precise initial estimates of the intrinsic and extrinsic parameters, therefore they propose an analytical solution to obtain good initial estimates, which are subsequently be refined using iterative least squares.

Velas et al. [6] have developed a method for joint lidar and monocular camera calibration using a target (plate) with four circular holes. The authors obtain coarse estimate based on projective geometry. After that, the pose estimate is refined by maximizing a cost function based on cross correlation between edges in images and point clouds. Guindel et al. [3] use the same calibration target to calibrate a lidar and a stereo camera. ICP is now used to minimize the error between the circle centers of lidar and stereo camera.

Recently, [17] proposed an online calibration method for lidar and camera using a Convolutional Neural Network to replace the traditional feature extraction, matching and registration steps. Several expert networks are developed to handle different calibration ranges (coarse to fine).

Several methods have been proposed to calibrate a radar with another sensor [5], [9], [21], [22]. To calibrate a radar and monocular camera, both [21] and [22] use a homography projection in a least squares formulation, without estimating the full 3D transformation. El Natour et al. [5] calibrate radar and monocular camera using a system of equations, with additional spherical and geometrical constraints.

Peršić et al. [9] use a triangular shaped styrofoam calibration target which also equips a metal trihedral corner reflector. The triangular shape can be detected in the lidar point cloud, while trihedral corner reflectors are commonly used targets for radars, because of their accurate localization and their distinct reflectivity: Radar Cross Section (RCS) value. Their calibration routine minimizes the reprojection error between lidar and radar. In addition, they propose a FOV refinement step using the RCS value of multiple targets.

B. Calibrating more than two sensors

While most works only focus on calibrating two sensors, to calibrate multiple sensors, one could minimize the effort with the discussed pair-wise calibration techniques by calibrating all sensor towards a single reference sensor, which we term *minimally connected pose estimation*. However, other configurations are suggested by studying related pose estimation problems in mobile robotics.

Simultaneous Localization and Mapping (SLAM) is the process of constructing a map of the environment and at the same time localizing a mobile sensor in the map [23]. When the system detects that it has returned to a previously visited location, it can use this 'loop closure' constraint to readjust all past poses along the trajectory. In extrinsic calibration, loop closure can similarly be added as a constraint too when optimizing all transformations between many sensors at once, which we will term *fully connected pose estimation*. Sim et al. [12] use it to calibrate a lidar and multiple cameras.

TABLE I: Related work on extrinsic sensor calibration. The columns with *L* ,*R, S* and *M* stand for lidar, radar, stereo camera and monocular camera, respectively. Symbols ✓ and ✗ indicate whether the method can calibrate a particular sensor. Symbol ∼ indicates that a stereo camera could be calibrated as two separate monocular camera, in principle, which is suboptimal. Furthermore, the abbreviations in column *Int./Ext.* refer to intrinsic calibration routine (Int.) and extrinsic calibration routine (Ext.). The column *SW* indicates if the software is open-source and available to the community.

	Configuration	Int./Ext.	Error Measure	L	R	S	M	SW	Toolbox name
Guindel et al. [3]	MCPE	Ext.	m	✓	✗	✓	✗	✓	velo2cam_calibration (ROS)
Geiger et al. [4]	MCPE	Int.&Ext.	m	✓	✗	∼	✓	∼	Online web toolbox
El Natour et al. [5]	MCPE	Ext.	m	✗	✓	∼	✓	✗	
Velas et al. [6]	MCPE	Ext.	pixels	✓	✗	∼	✓	✓	but_calibration_camera_velodyne (ROS)
Alismail et al. [7]	MCPE	Int.&Ext.	m	✓	✗	∼	✓	✓	calidar (MATLAB)
Zhang & Pless [8]	MCPE	Ext.	m, pixels	✓	✗	∼	✓	✓	RADLOCC (MATLAB)
Peršić et al. [9]	MCPE	Ext.	m	✓	✓	✗	✗	✗	
Dhall et al. [10]	MCPE	Ext.	m	✓	✗	∼	✓	✓	lidar_camera_calibration (ROS)
Mirzaei et al. [11]	MCPE	Int.&Ext.	m	✓	✗	∼	✓	✗	
Sim et al. [12]	MCPE/FCPE	Ext.	pixels	✓	✗	∼	✓	✗	
Proposed	MCPE/FCPE/PSE	Ext.	m, likelihood	✓	✓	✓	✓	✓	

Visual Odometry involves estimating the ego-motion from visual features extracted from consecutive images, after that bundle adjustment can be used for pose estimation refinement [24]. Since multiple poses are considered, bundle adjustment fuses all information in a probabilistic model which explicitly represents the 3D structure of the features, and optimizes a maximum likelihood criterion [13]. Such a formulation could also benefit joint calibration of more than two sensors, which was not explored for most available methods (see table I).

C. Contributions

In contrast to the discussed related work, our work provides the following contributions. First, three extrinsic calibration configurations to jointly calibrate lidar, camera and radar are investigated. We study the three configurations, required number of calibration board locations, and choice for the reference sensor using a real multi-modal sensor setup. Second, we propose a calibration target design that is detectable by lidar, camera and radar. Third, we provide an open-source extrinsic calibration tool for these sensors, with bindings to Robot Operating System (ROS)[1].

III. Proposed Approach

This section details our novel tool, which uniquely calibrates lidar, camera and radar jointly, see table I. It is implemented as an open-source tool with bindings to the ROS middleware, and which includes a tool that can update Unified Robot Description Format (URDF), to facilitate extensibility and application on real robotic platforms.

First, the design of the calibration board is described, then the detectors that extract features from the raw sensor data are presented. Then, a mathematical description of the calibration of two sensors is given, which we then extend to the joint calibration of more sensors.

a) Calibration target design: In order to jointly calibrate multiple sensors, the calibration target should be detectable in all relevant modalities. For lidar and camera, edges and corners are features which can be detected accurately and robustly. However, rectangular shaped objects are difficult to localise in lidar as a nearly horizontal edge might not intersect with any of the lidar scan planes [9]. We follow [3], [6] and use circular shapes, which can be accurately detected when intersecting with few lidar beams. Our proposed calibration target design has four circular holes and additionally contains a single metal trihedral corner reflector in the center at the back of the board to provide strong radar reflections as well. Furthermore, our calibration board is made from styrofoam to not affect the detectability of the corner reflector [9]. Figure 3 illustrates 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 centres $a_2 = 24$ cm. The reflector is at $a_3 = 10.5$ cm from the front.[2]

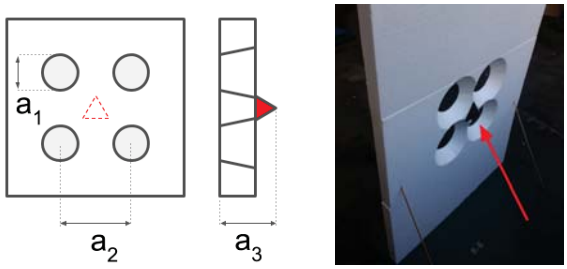


Fig. 3: From left to right, front view drawing, side view drawing, and an image of the back of the target. The trihedral corner reflector is indicated in red (triangle and arrow).

b) Detection of calibration target: Both the lidar detector and the stereo detector are an adapted version of the detector of [3]. Both camera and lidar detectors return the 3D locations of the four circle centers. Having four feature points facilitates identifying incorrect detections, since the ratio between the maximum distance (diagonal) and the minimum distance (side of square) should equal $\sqrt{2}$. Outlier boards that deviate from this ratio are discarded.

The radar returns detections in a 2D plane and generates for each reflection a measurement in polar coordinates and a RCS value. Of all detections that are within the expected RCS range, the closest radar detection to the car is taken.

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

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

Finally, our tool also includes a monocular camera detector however our experiments will focus on stereo, lidar and radar.

c) Calibration procedures: To formalize the calibration of more than two sensors, we will first introduce extrinsic calibration for two sensors.

Consider two sensors, indicated by numbers 1 and 2. After placing the target calibration board at K different locations in front of the sensors, each detector provide K detections $\boldsymbol{y}^1 = \{\boldsymbol{y}_1^1, \cdots, \boldsymbol{y}_K^1\}$ and $\boldsymbol{y}^2 = \{\boldsymbol{y}_1^2, \cdots, \boldsymbol{y}_K^2\}$ of the targets, relative to coordinate frames of sensor 1 and 2 respectively. For camera and lidar, each detection consists of four 3D coordinates of the circle centers, i.e. $\boldsymbol{y}_k = (y_{k(1)}, \cdots, y_{k(4)})$. A radar detection $\boldsymbol{y}_k = (y_{k(1)})$ has only a single location $y_{k(1)}$ for the detected trihedral corner reflector, which is expressed in 2D Euclidean coordinates.

Since each sensor has a different Field of View (FOV), the calibration target may not always be detectable by all. Variables μ_k^i indicates if calibration board k was detected for sensor i, thus $\mu_k^i = 0$ if the target was not found or discarded, and 1 otherwise. Extrinsic calibration between the two sensors aims to estimate the relative rigid transformation $T^{1,2}$ which projects a point from sensor 1 onto the coordinate frame of sensor 2. The rigid transformation consists of a 3×3 rotation matrix R and 3D translation t, which is described as a 4×4 matrix for homogeneous coordinates,

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

Similarly, we will assume homogeneous representations in $\boldsymbol{y}^1$ and $\boldsymbol{y}^2$, hence each 3D point (x, y, z) is represented as an augmented 4D vector $(x, y, z, 1)$. To parameterize the 6 degrees of freedom of transformation $T^{1,2}$, we use parameter vector $\theta^{1,2} = (t_x, t_y, t_z, v_x \cdot \alpha, v_y \cdot \alpha, v_z \cdot \alpha)$. The rotation is here 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 transformation error between camera and/or lidar detections is the total squared Euclidean distance of the four detected circle centers,

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

If one of the sensors is a radar, however, a different error term is used. Let $\boldsymbol{y}_k^R$ represents the radar measurement of target k, then the squared Euclidean error now becomes

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

Here $g(\boldsymbol{y}_k)$ computes the expected 3D position of the trihedral corner reflector given the four 3D circle positions in detection $\boldsymbol{y}_k$, using the known calibration target's geometry. The function $p(q_k)$ first converts 3D Euclidean point q_k to spherical coordinates (r_k, ϕ_k, ψ_k), then disregards the elevation angle ψ_k, and converts the remaining polar coordinates (r_k, ϕ_k) to their 2D Euclidean equivalent.

Additionally, we enforce the constraint that the projected 3D points lie within radar Field of View (FOV). For instance, we add constraints that the elevation angles ψ_k for all calibration board locations k are within the maximum view angle ψ_{max} of the radar,

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

Pose estimation can now be formulated as an optimization problem to find the optimal transformation which minimizes the total error $f(\theta^{1,2})$ between all K calibration targets,

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

The indicator variables $\mu_k^2 \cdot \mu_k^1$ ensure that only terms are included where the target was detected in both sensors. The optimal calibration parameters are thus found by minimizing the error criterion $f(\theta)$, which could be subject to zero or more (in)equality constraints (e.g. equation (4)).

Our tool uses Sequential Least SQuares Programming (SLSQP) from the *SciPy* library [25] for optimization. An initial solution is obtained by using the Kabsch algorithm. For radar, it is assumed that detections lie on the radar plane (zero elevation angle) in order to find an initial pose estimate.

d) Joint calibration with more than two sensors: Generally, a sensor setup can contain more than two sensors. We now consider three possibilities to adapt the extrinsic calibration procedure to optimize $\boldsymbol{\theta}$, the set of all transformation parameters for N sensors. The first two configurations can be considered different generalizations of the standard pairwise calibration discussed before. The third configuration is based on explicit structure estimation, similar to pose estimation found in graph SLAM [26] and bundle adjustment [27]. Each configuration will now be discussed in detail.

1) Minimally connected pose estimation (MCPE): First, sensors can be calibrated pairwise with respect to a selected 'reference' sensor. This results in a *minimally connected* graph, where the edges describe the transformation of the 'reference' to sensor, see figure 2a. Without loss of generality, let's assume that the first sensor is selected as reference. The optimization function is now reformulated as

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

Note that transformations between any non-reference sensors i, j, follow from joining the transformations on the unique path of this graph, i.e. $T^{i,j} = T^{1,j} \cdot (T^{1,i})^{-1}$.

2) Fully connected pose estimation (FCPE): Next, we consider optimizing transformations between all sensors at once, without a special reference sensor. As shown in figure 2b, this results in optimizing terms in a fully connected graph, akin to a loop closure optimization in SLAM methods. 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 estimated, resulting in the following total error function,

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

In addition, the closed loop constraint is introduced to ensure that all loops l equal the identity matrix, namely

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

where s_l equals the number of sensor in this loop l. By adding more error terms, the optimization is potentially more robust against noisy observations from one reference sensor. The downside is that the number of error terms increases quadratically with the number of sensors. Furthermore, additional loop constraints must be added as well.

3) Pose and structure estimation (PSE): The third and final considered configuration is pose and structure estimation, which has similarities to bundle adjustment. This configuration explicitly estimates the calibration board poses and the observation noise of each sensor. The objective is to estimate both the unknown structure $M = (m_1, \cdots, m_K)$ of the true target poses in a fixed coordinate frame, and the transformation $T^{M,i}$ from the fixed frame to each sensor i, see figure 2c. 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,

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

As a result from this formulation, we use the squared Mahalanobis distance instead of the squared Euclidean distance, which for vectors a and b, and covariance Σ is written as

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

For pose estimation, we first initialize all Σ^i as identity, and jointly optimize the transformations and structure,

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

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

The result is used to re-estimate the diagonal elements of the noise covariances, after which the optimization of $f(\boldsymbol{\theta}, M)$ is repeated. This process is iterated until all variances have converged. Note that one transformation $T^{M,i}$ must be fixed in order to avoid that a solution is not uniquely determined.

The potential benefit is that this is a homogeneous error metric. For every sensor, we can express the error as a negative log likelihood whereas combining squared errors might result in heterogeneous error functions (pixel versus Euclidean). Another benefit of such a probabilistic formulation is that prior knowledge on board and sensor poses could be included, we have not pursued this direction here. The disadvantages are that the optimization is more complex and that the loop closure constraint is not explicitly enforced.

IV. Experiments

For our experiments, we use a sensor setup that is mounted on a Toyota Prius. It consists of: a Velodyne HDL-64E lidar (on roof), a Continental ARS430 radar (behind bumper), and a stereo camera 2× UI-3060CP Rev. 2 (behind windscreen) using dense Semi-Global Matching [28]. With these sensors, we have recorded a dataset with sensor measurements of 29 calibration board locations that are located in the FOV of all sensors and in the working range of 5 m from the car.

We report the calibration performance for the three sensor pairs: lidar to camera ($l2c$), lidar to radar ($l2r$), camera to radar ($c2r$). The used measure is the Root Mean Squared Error (RMSE) on all available detections of all 29 calibration boards when applying the estimated transformation.

a) Choice of MCPE reference sensor: First, we assess if the choice for the reference sensor in the configuration MCPE affects the calibration performance. Table II lists the performance for all three reference sensors, as well as the other two configurations when using 5 random calibration board locations. All reference sensor choices show similar performance, however selecting the radar as reference sensors lacks optimization of the only full 3D transformation of $l2c$, resulting in less accurate results in some cases. We will use lidar as MCPE reference sensor in the other experiments. Note that $l2c$ and $l2r$ generally obtain the most accurate results, and that both FCPE and PSE perform better than MCPE for the $l2r$ and $c2r$ transforms.

TABLE II: Mean and standard deviation of the RMSE [m] for 100 combinations of 5 calibration board locations.

	RMSE l2c [m]	RMSE l2r [m]	RMSE c2r [m]	Summed RMSE
MCPE(camera)	0.018±0.002	0.023±0.003	0.031±0.004	0.071
MCPE(lidar)	0.018±0.002	0.022±0.003	0.030±0.004	0.070
MCPE(radar)	0.018±0.002	0.022±0.003	0.031±0.004	0.071
FCPE	0.018±0.002	0.020±0.004	0.027±0.005	0.065
PSE	0.018±0.003	0.019±0.003	0.025±0.003	0.062

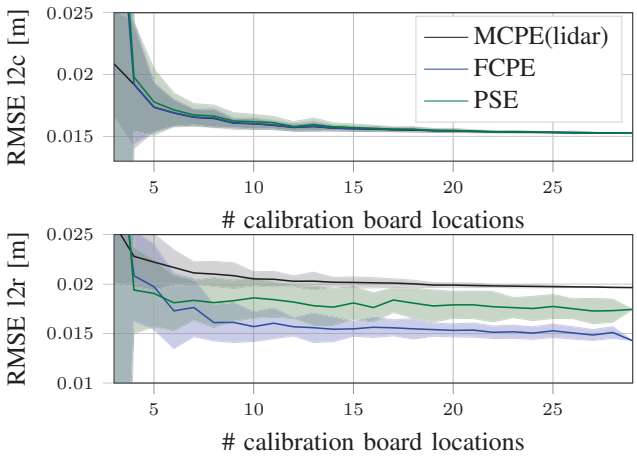


Fig. 4: The mean and standard deviation of the RMSE on 50 board locations for varying K. The top figure shows the RMSE for $l2c$ and the bottom figure the RMSE for $l2r$.

b) Comparison to baseline method: Next, we assess the benefit of our tool handling multiple targets by comparing our tool with the publicly available calibration method of Guindel et al. [3] for lidar to stereo camera calibration only. Table III shows that no differences were found for single board calibration with [3] and our MCPE implementation,

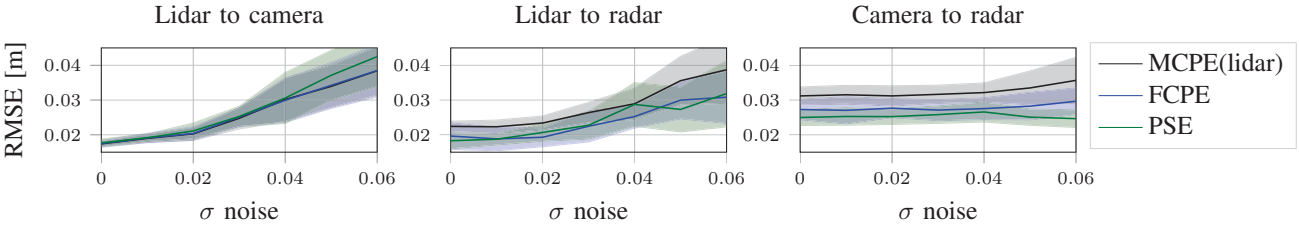


Fig. 5: RMSE error as function of observation noise added to the lidar observations with standard deviation as defined on the x-axis. The median and the median absolute deviation are plotted based on 50 combinations of 5 calibration board locations.

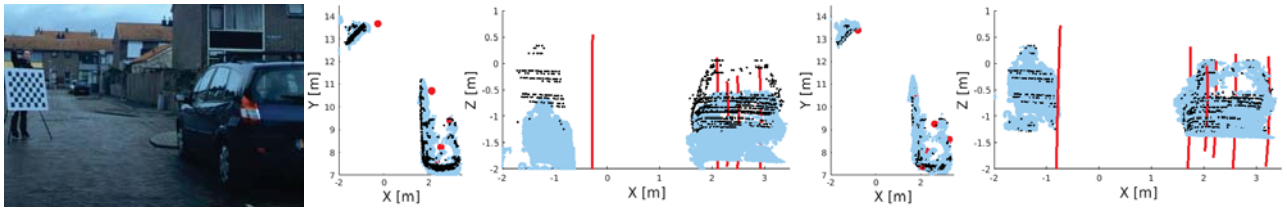


Fig. 6: The left image shows the recorded scene to test the calibration. There is a parked car $\sim 6m$ in front of the sensor setup, and a person with a checkerboard at $\sim 13m$. The middle two plots show the lidar (black) and stereo (blue) point cloud, and radar detections (red) before extrinsic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

TABLE III: Single board versus multiple board methods.

lidar to stereo method	# boards	RMSE [m]
Guindel et al. [3]	single board	0.0393 ± 0.0104
MCPE	single board	0.0393 ± 0.0104
MCPE	all boards (29)	0.0153
FCPE	all boards (29)	0.0153
PSE	all boards (29)	0.0153

which we repeated for all 29 board locations. For single board methods, the mean and standard deviation of the RMSE are reported. Furthermore, the table shows that using more board locations can yield a $2\times$ reduction of the RMSE.

c) Required number of calibration board locations: We follow up the previous experiment by investigating how the number of available board locations affects the result. All our methods are tested on varying the number of locations K, where for each value of K we use 50 fixed sets of K randomly selected locations. Figure 4 shows how the mean and standard deviation of the RMSE over all 50 sets changes with K. The figure shows that when 10 boards are used, the errors for FCPE and PSE have converged to $\leq$ 2 cm for both $l2c$ and $l2r$ transforms. MCPE performs the worst for the RMSE of $l2r$, whereas FCPE performs the best with a RMSE of less than 1.5 cm when using all 29 board locations. We have observed that PSE is sometimes less robust.

d) Sensitivity to observation noise: For each method, we also study the sensitivity to sensor observation noise by adding varying amounts of zero-mean Gaussian noise $\mathcal{N}(0, \sigma^2 I_3)$ to the 3D measurements of the lidar detections. Figure 5 shows the median and median absolute deviation of the RMSE for various values of σ. It can be seen that all errors related to lidar increase when the noise levels increase. Unlike MCPE, both FCPE and PSE are robust against additive Gaussian noise for the $c2r$ error. Since the $c2r$ link is not present in MCPE, its performance is affected more when noise levels increase.

e) Qualitative results: Finally, we validated our tool in an outdoor scene. Initial sensor poses were obtained by manually adjusting the relative sensor poses, and then extrinsic calibration was performed using the PSE configuration. Figure 6 shows the lidar and stereo point clouds, as well as the radar detections, before and after calibration. Initially, the lidar and stereo point cloud did not align, especially in the Z direction, and the left-most radar detection did not correspond to points in the lidar or stereo point cloud. After calibration, the two point clouds closely match, and the radar detections also coincide with the objects.

V. Conclusion

We presented an open-source extrinsic calibration tool for lidar, camera and radar, and proposed three configurations to estimate the sensor poses from simultaneous detections of multiple calibration board locations. Experiments on a setup with all sensing modalities show that all configurations can provide good calibration results. Furthermore, the results with five calibration board locations show that the expected RMSE is approximately 2 cm for lidar to camera and lidar to radar, and approximately 2.5 cm for camera to radar. When using more than five board locations, *fully connected pose estimation* shows the best performance. Future work involves investigating the effect of more than three sensors on the calibration performance.

Acknowledgement

The work is supported by NWO TTW under the project STW#13434 Standardized Self-Diagnostic Sensing Systems for Highly Automated Driving.

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 Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1773–1795, 2013.

[2] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 7, pp. 1239–1258, 2010.

[3] C. Guindel, J. Beltrán, D. Martín, and F. García, "Automatic extrinsic calibration for lidar-stereo vehicle sensor setups," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, pp. 1–6.

[4] A. Geiger, F. Moosmann, Ö. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 3936–3943.

[5] 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. 25 937–25 967, 2015.

[6] M. Velas, M. Španěl, Z. Materna, and A. Herout, "Calibration of RGB camera with velodyne lidar," *WSCG 2014: communication papers proceedings*, 2014.

[7] H. Alismail, L. D. Baker, and B. Browning, "Automatic calibration of a range sensor and camera system," in *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*. IEEE, 2012, pp. 286–292.

[8] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, vol. 3. IEEE, 2004, pp. 2301–2306.

[9] J. Peršić, I. Marković, and I. Petrović, "Extrinsic 6DoF calibration of 3D lidar and radar," in *2017 European Conference on Mobile Robots (ECMR)*. IEEE, 2017, pp. 1–6.

[10] A. Dhall, K. Chelani, V. Radhakrishnan, and K. M. Krishna, "LiDAR-Camera Calibration using 3D-3D Point correspondences," *ArXiv e-prints*, May 2017.

[11] F. M. Mirzaei, D. G. Kottas, and S. I. Roumeliotis, "3D LIDAR–camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization," *The International Journal of Robotics Research*, vol. 31, no. 4, pp. 452–467, 2012.

[12] S. Sim, J. Sock, and K. Kwak, "Indirect correspondence-based robust extrinsic calibration of lidar and camera," *Sensors*, vol. 16, no. 6, p. 933, 2016.

[13] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[14] R. Szeliski, *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

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

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

[17] N. Schneider, F. Piewak, C. Stiller, and U. Franke, "RegNet: Multimodal sensor registration using deep neural networks," in *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2017, pp. 1803–1810.

[18] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[19] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, vol. 32, no. 5, pp. 922–923, 1976.

[20] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, Feb 1992.

[21] S. Sugimoto, H. Tateda, H. Takahashi, and M. Okutomi, "Obstacle detection using millimeter-wave radar and its visualization on image sequence," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3. IEEE, 2004, pp. 342–345.

[22] 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.

[23] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[24] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *IEEE Robotics & Automation Magazine*, vol. 18, no. 4, pp. 80–92, 2011.

[25] E. Jones, T. Oliphant, P. Peterson, *et al.*, "SciPy: Open source scientific tools for Python," 2001–, [Online; accessed September 2018]. [Online]. Available: http://www.scipy.org/

[26] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press, 2005.

[27] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment - a modern synthesis," in *International workshop on vision algorithms*. Springer, 1999, pp. 298–372.

[28] D. Hernandez-Juarez, A. Chacón, A. Espinosa, D. Vázquez, J. C. Moure, and A. M. López, "Embedded real-time stereo estimation via semi-global matching on the GPU," in *International Conference on Computational Science (ICCS) 2016, 6-8 June 2016, San Diego, California, USA*, 2016, pp. 143–153.