2021 IEEE International Conference on Robotics and Automation (ICRA 2021)
May 31 - June 4, 2021, Xi'an, China

A Continuous-Time Approach for 3D Radar-to-Camera Extrinsic Calibration

Emmett Wise[1], Juraj Peršić[2], Christopher Grebe[1], Ivan Petrović[2], and Jonathan Kelly[1,†]

***Abstract*— Reliable operation in inclement weather is essential to the deployment of safe autonomous vehicles (AVs). Robustness and reliability can be achieved by fusing data from the standard AV sensor suite (i.e., lidars, cameras) with *weather robust* sensors, such as millimetre-wavelength radar. Critically, accurate sensor data fusion requires knowledge of the rigid-body transform between sensor pairs, which can be determined through the process of extrinsic calibration. A number of extrinsic calibration algorithms have been designed for 2D (planar) radar sensors—however, recently-developed, low-cost 3D millimetre-wavelength radars are set to displace their 2D counterparts in many applications. In this paper, we present a continuous-time 3D radar-to-camera extrinsic calibration algorithm that utilizes radar velocity measurements and, unlike the majority of existing techniques, does not require specialized radar retroreflectors to be present in the environment. We derive the observability properties of our formulation and demonstrate the efficacy of our algorithm through synthetic and real-world experiments.**

I. Introduction

Safety is a paramount concern for autonomous vehicles (AVs) operating in human-centric environments (e.g., self-driving cars travelling on city streets). To reduce the risk of failure and improve robustness, most AVs fuse data from multiple sensors on board. The standard AV sensor suite typically includes cameras and lidar units; while these sensors are able to provide a high degree of situational awareness, they may fail to work reliably in inclement weather (e.g., heavy rain or snowfall). In turn, many AV sensor platforms incorporate 2D (planar) millimetre-wavelength radar units that are *weather robust*—radar measurements are relatively immune to interference caused by precipitation, for example.

All radar sensors operate on the same basic principle: a low-frequency electromagnetic (EM) pulse is emitted from the radar antenna, reflects off of radar-opaque targets in the environment, and returns to the sensor. By measuring the time of flight and phase of the return pulse, the radar is able to determine the azimuth, range, range-rate (velocity in the radial direction), and cross-section (reflectivity) of targets. Low-frequency EM waves are able to pass through rain, snow, and other obscurants [1]. Although 2D radar has proven useful for many AV applications, the lack of complete 3D information limits its utility in many cases.

[1]Emmett Wise, Christopher Grebe, and Jonathan Kelly are with the Space & Terrestrial Autonomous Robotics Systems (STARS) Laboratory at the University of Toronto Institute for Aerospace Studies, Toronto, Canada. <firstname>.<lastname>@robotics.utias.utoronto.ca

[2]Juraj Peršić and Ivan Petrović are with the Laboratory for Autonomous Systems and Mobile Robotics, University of Zagreb Faculty of Electrical Engineering and Computing, Croatia. <firstname>.<lastname>@fer.hr

†Jonathan Kelly is a Vector Institute Faculty Affiliate. This research was supported in part by the Canada Research Chairs program.

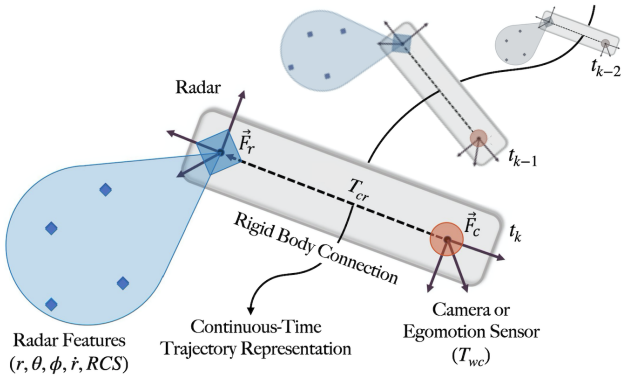

Fig. 1. Depiction of the calibration problem. The radar measures the range, azimuth, elevation, range-rate, and reflectivity of objects in the environment. The camera (or egomotion sensor) measures its own pose change relative to a fixed reference frame. Our goal is to recover the rigid-body transform $\mathbf{T}_{cr}$ between the radar unit and the camera.

More recently, low-cost 3D radar sensors, such as the Texas Instruments AWR1843BOOST, have become available. Because of the additional information contained in 3D radar measurements (i.e., elevation), 3D radars are poised to replace 2D sensors in AV systems and in other applications. To properly fuse 3D radar data with measurements from other AV sensors, however, knowledge of the rigid-body transform between the radar and the other sensors is required. The process of determining the transform is known as extrinsic calibration. Often, extrinsic calibration is performed prior to deployment, in a laboratory or factory setting; the transform parameters are prone to change, however, due to material fatigue or user modifications. Consequently, there is a need for methods to estimate the extrinsic calibration in the field.

Radar extrinsic calibration is challenging for several reasons. First, most radar measurement models assume that the EM pulse is reflected by one surface only. In reality, there are often multipath reflections from several different surfaces. These multipath reflections create measurement outliers that can obscure or 'drown out' the true reflection from a target. Second, raw radar measurements have substantial jitter, which reduces measurement precision. Finally, a radar pulse is a wave, and hence the exact point of reflection from a target can be ambiguous and/or inconsistent [2]. The low precision and high outlier rate of radar measurements can degrade estimates of the extrinsic calibration. To mitigate some of these issues, many existing calibration algorithms rely on specialized radar retroreflectors that are placed strategically in the environment. Although this approach improves calibration, specialized retroreflectors are rarely available in the field during regular operation.

We overcome the challenges of radar extrinsic calibration by relying on the *motion* of the sensor platform rather than on specific scene structure (see Fig. 1). Work by Stahoviak has shown that the velocity of a 3D millimetre-wavelength (hereafter, mm-wave) radar sensor can be determined directly and without knowledge of the environment [3]. By relying on velocity information provided by the 3D radar, instead of attempting to localize and track specific targets, we avoid many of the issues caused by noise, outliers, and jitter. We focus on radar-to-camera extrinsic calibration—however, the method we describe is applicable to any complementary sensor that is able to estimate its egomotion (e.g., 3D lidar, GNSS/INS sensors, etc.). We require only enough information for egomotion estimation and sufficient excitation of the system (see Section IV-B). In this paper we:

1) prove that extrinsic calibration for a 3D radar-camera pair is observable given sufficient excitation of the system;
2) describe the required motions necessary for proper calibration;
3) develop a continuous-time batch radar-to-monocular camera extrinsic calibration algorithm; and
4) verify the performance of our algorithm on synthetic data and through extensive real-world experiments.

We provide one of the first methods for estimating the extrinsic calibration parameters between a 3D mm-wave radar and monocular camera without the use of radar retroreflectors. Although our goal is to build weather-robust navigation platforms, we focus on calibration under nominal conditions in the field (i.e., without adverse weather), since this is already a very difficult problem.

II. Related Work

A variety of mm-wave radar extrinsic calibration algorithms exist, which can roughly be grouped according to the sensor pair involved and the specific degrees of freedom that are calibrated. Early extrinsic calibration algorithms for radar-camera sensor pairs considered 2D radar units only, either ignoring the 3D nature of radar measurements or constraining the positions of any retroreflectors to the radar measurement plane [4]–[7]. These algorithms operate by estimating the homography between the camera image plane and the radar measurement plane. Sugimoto et al. note in [4] that 2D radar units typically measure a maximum return when a retroreflector lies on the plane of zero elevation in the radar reference frame; the return intensity decreases for reflectors that lie above or below this plane. The approach in [4] filters returns by intensity to ensure that only targets in the plane at zero elevation (relative to the radar frame) are used as part of the calibration process.

More recent algorithms estimate the rigid sensor-to-sensor transform by minimizing a 'reprojection error': this is the error in the alignment of identifiable environmental structures or objects that appear within the fields of view of both sensors. Kim et al. [8] align hybrid visual-radar targets that can be easily identified in the camera and radar data, but assume that the radar measurements are constrained to the zero-elevation plane.

The zero-elevation plane constraint is relaxed for certain 'reprojection error' algorithms. El Natour et al. estimate the radar-to-camera transform by intersecting backprojected camera rays with the 'arcs' in 3D along which radar measurements must lie [9]. Domhof et al. rely on a known visual target structure to convert camera measurements into 'pseudo-radar' measurements. The transform that best aligns the radar and pseudo-radar measurements then defines the extrinsic calibration [10]. Peršić et al. [11] improve upon these methods by resolving the elevation ambiguity using target reflection intensity as a pseudo-measurement of the elevation angle. Peršić et al. [11] also extend their approach to include 2D radar-to-lidar calibration. The reprojection and homography methods are summarized and compared by Oh et al. in [12], where the authors conclude that the homography and reprojection methods have similar accuracy.

All of the algorithms described above require specialized retroreflective radar targets, but a small number of 'targetless' or target-free extrinsic calibration algorithms for 2D mm-wave radar also exist. Schöller et al. [13] use end-to-end deep learning to estimate the extrinsic rotation parameters that align vehicles (i.e., automobiles) detected in radar measurements and camera images. However, the algorithm requires an external measurement of the translation parameters. Peršić et al. [14] perform target-free, online pairwise extrinsic calibration of 2D radars, cameras, and lidar sensors by estimating the transform that aligns moving object trajectories. This method assumes a priori knowledge of the translation parameters and only estimates yaw between the radar-camera and radar-lidar pairs.

Similar to our approach, Kellner et al. [15] use radar velocity measurements to estimate the yaw angle between a 2D radar sensor and a vehicle-mounted gyroscope, by relating the angular velocity of the gyroscope to the lateral velocity of the radar. This technique also requires a priori knowledge of the translation between the sensors.

In summary, the mm-wave radar calibration algorithms developed to date are generally limited by hardware constraints (i.e., an inability to resolve elevation reliably) or the need for specialized retroreflective targets, or suffer from high calibration parameter uncertainty due to a lack of true 3D information. We take advantage of the available elevation data in 3D radar measurements to estimate the instantaneous (3D) velocity of the radar unit. These data, in combination with pose estimates from a camera (or other egomotion sensors), allow us to determine the full sensor-to-sensor rigid-body transform without the need for specialized targets.

III. Problem Formulation

A. Notation

Latin and Greek letters (e.g., a and α) represent scalar variables, while boldface lower and upper case letters (e.g., $\mathbf{x}$ and $\mathbf{\Theta}$) represent vectors and matrices, respectively. A parenthesized superscript pair, for example, $\mathbf{A}^{(i,j)}$, indicates

the ith row and the jth column of the matrix $\mathbf{A}$. A three-dimensional reference frame is designated by $\underrightarrow{\mathcal{F}}$. The translation vector from point a (often a reference frame origin) to b, expressed in $\underrightarrow{\mathcal{F}}_a$, is denoted by $\mathbf{r}_a^{ba}$. The translational velocity of point b relative to point a, expressed in $\underrightarrow{\mathcal{F}}_c$, is denoted by $\mathbf{v}_c^{ba}$. The angular velocity of frame $\underrightarrow{\mathcal{F}}_a$ relative to an inertial frame, expressed in $\underrightarrow{\mathcal{F}}_a$, is denoted by $\boldsymbol{\omega}_a$.

We denote rotation matrices by $\mathbf{R}$; for example, $\mathbf{R}_{ab} \in \mathrm{SO}(3)$ defines the rotation from $\underrightarrow{\mathcal{F}}_b$ to $\underrightarrow{\mathcal{F}}_a$. We reserve $\mathbf{T}$ for $\mathrm{SE}(3)$ transform matrices; for example, $\mathbf{T}_{ab}$ is the homogeneous matrix that defines the rigid-body transform from frame $\underrightarrow{\mathcal{F}}_b$ to $\underrightarrow{\mathcal{F}}_a$. These transforms are constructed using the split representation of $\mathrm{SE}(3)$. For example, the transform from frame $\underrightarrow{\mathcal{F}}_b$ to $\underrightarrow{\mathcal{F}}_a$ at time t is,

$$\mathbf{T}_{ab}(t) = \begin{bmatrix} \mathbf{R}_{ab}(t) & \mathbf{r}_a^{ba}(t) \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{1}$$

where the transform is split into a rotation matrix, $\mathbf{R}_{ab}(t) \in \mathrm{SO}(3)$, and translation vector, $\mathbf{r}_a^{ba}(t) \in \mathbb{R}^3$. The unary operator $^\wedge$ acts on $\mathbf{r} \in \mathbb{R}^3$ to produce a skew-symmetric matrix such that $\mathbf{r}^\wedge\mathbf{s}$ is equivalent to the cross product $\mathbf{r} \times \mathbf{s}$.

B. Sensor Measurements

We consider three reference frames: frame $\underrightarrow{\mathcal{F}}_w$ is an (approximate) inertial frame attached to the surface of the Earth, while $\underrightarrow{\mathcal{F}}_r$ is the reference frame of the radar sensor, and $\underrightarrow{\mathcal{F}}_c$ is the reference frame of the camera (or other egomotion sensor). The radar unit measures the velocity of the sensor in $\underrightarrow{\mathcal{F}}_r$ relative to $\underrightarrow{\mathcal{F}}_w$, expressed in $\underrightarrow{\mathcal{F}}_r$ at an instant in time, t,

$$\mathbf{v}_r^{rw}(t) = \mathbf{R}_{wr}(t)^T \frac{\partial \mathbf{r}_w^{rw}(t)}{\partial t}, \tag{2}$$

where we use the partial derivative notation to indicate that the radar position also depends upon the parameters of our B-spline trajectory representation (see Section III-C).

Assuming that a series of three or more (known) 3D landmarks are visible in frame $\underrightarrow{\mathcal{F}}_w$, the camera is able to measure its pose at time t relative to $\underrightarrow{\mathcal{F}}_w$,

$$\mathbf{T}_{cw}(t) = \mathbf{T}_{cr}\mathbf{T}_{wr}^{-1}(t), \tag{3}$$

where $\mathbf{T}_{wr}(t)$ is the homogeneous pose matrix of the radar in the inertial frame at time t and $\mathbf{T}_{cr}$ is the homogeneous matrix that defines the (constant but unknown) radar-to-camera transform. If the metric positions of the landmarks are not known, the camera translation can only be determined up to an unknown scale factor.

C. Continuous-Time Trajectory Representation

We use a continuous-time representation of the sensor platform trajectory in our problem formulation. The continuous-time representation is advantageous because it allows measurements to be made at arbitrary time instants; since the radar and the camera operate at different rates and are not hardware synchronized, the relationship between their measurement times is not fixed. There are multiple possible ways to parameterize trajectories in continuous time [16]–[18]. We choose the cumulative B-spline representation on Lie groups developed by Sommer et al. in [16]. Below, we very briefly review this representation, and refer the reader to [16] for more details.

B-splines are functions of one continuous parameter (e.g time) and a finite set of control points (or *knots*); for brevity, we restrict our example here to points $\{\mathbf{p}_0, \ldots, \mathbf{p}_N \mid \mathbf{p}_i \in \mathbb{R}^d\}$. The order k of the spline determines the number of control points that are required to evaluate the spline at time t. In a uniformly spaced B-spline, each control point is assigned a time $t_i = t_0 + i\Delta t$, where t_0 is the start of the spline and Δt is the time between control points. Given a B-spline of length N and order k, the end of the spline is t_{N-k+1}.

Given a time t, a normalized time $u = \frac{t-t_i}{t_{i+1}-t_i}$ can be defined, where t_i is the time assigned to control point $\mathbf{p}_i$ and $t_i \leq t < t_{i+1}$. The B-spline function evaluated at normalized time u is

$$\mathbf{p}(u) = \begin{bmatrix} \mathbf{p}_i & \mathbf{d}_1^i & \ldots & \mathbf{d}_{k-1}^i \end{bmatrix} \tilde{\mathbf{M}}_k \mathbf{u}, \tag{4}$$

where $\mathbf{u}^T = [1 \; u \; u^2 \; \ldots \; u^{k-1}]$ and $\mathbf{d}_j^i = \mathbf{p}_{i+j} - \mathbf{p}_{i+j-1}$. The matrix $\tilde{\mathbf{M}}_k$ is a $k \times k$ *mixing matrix*. The elements of the mixing matrix are a function of the spline order k and are defined by

$$\tilde{m}_k^{(a,n)} = \sum_{s=a}^{k-1} m_k^{(s,n)}, \tag{5}$$

$$m_k^{(s,n)} = \frac{C_{k-1}^n}{(k-1)!} \sum_{l=s}^{k-1} (-1)^{l-s} C_k^{l-s} (k-1-l)^{k-1-n} \quad a, s, n \in \{0, \ldots, k-1\}. \tag{6}$$

The scalar $C_j^i = \frac{j!}{i!(j-i)!}$ is a binomial coefficient. This B-splines definition can be simplified by defining $\lambda_j(u) = \tilde{\mathbf{M}}_k \mathbf{u}$, which results in

$$\mathbf{p}(u) = \mathbf{p}_i + \sum_{j=1}^{k-1} \lambda_j(u) \mathbf{d}_j^i. \tag{7}$$

This B-spline representation is a convenient way to describe smooth rigid-body trajectories in continuous time. Our development above is for splines on a vector space, but B-splines can also be defined over Lie groups, including the group $\mathrm{SO}(3)$ of rotations,

$$\mathbf{R}(u) = \mathbf{R}_i \prod_{j=1}^{k-1} \exp(\lambda_j(u) \boldsymbol{\phi}_j^i), \tag{8}$$

where $\mathbf{R}_i$ is a control point of the rotation spline and $\boldsymbol{\phi}_j^i = \log(\mathbf{R}_{i+j-1}^T \mathbf{R}_{i+j})$. The operators $\exp$ and $\log$ map from the Lie algebra $\mathfrak{so}(3)$ to $\mathrm{SO}(3)$ and vice versa, respectively [18].

D. Optimization Problem

The error equation for the radar velocity is

$$\mathbf{e}_v(t) = \mathbf{v}_r^{rw}(t) - \mathbf{R}_{wr}(t)^T \frac{\partial \mathbf{r}_w^{rw}(t)}{\partial t} + \mathbf{n}_v, \quad \mathbf{n}_v \sim \mathcal{N}(0, \boldsymbol{\Sigma}_v(t)), \tag{9}$$

where $\mathbf{R}_{wr}(t)$ and $\mathbf{r}_w^{rw}(t)$ are the split spline representation of $\mathbf{T}_{wr}(t)$ with control points $\{\mathbf{R}_0, \ldots, \mathbf{R}_N \mid \mathbf{R}_i \in \mathrm{SO}(3)\}$ and $\{\mathbf{p}_0, \ldots, \mathbf{p}_N \mid \mathbf{p}_i \in \mathbb{R}^3\}$. The vector $\mathbf{v}_r^{rw}(t)$ is the measured radar velocity at time t. The error equation for the camera measurements is

$$\mathbf{T}_{err}(t) = \mathbf{T}_{cw}(t)\mathbf{T}_{wr}(t)\mathbf{T}_{cr}^{-1} \tag{10}$$

$$\mathbf{e}_p(t) = \begin{bmatrix}\mathbf{r}_{err}(t)\\ \boldsymbol{\phi}_{err}(t)\end{bmatrix} + \mathbf{n}_p,\ \mathbf{n}_p \sim \mathcal{N}(0, \boldsymbol{\Sigma}_p(t)) \tag{11}$$

$$\boldsymbol{\phi}_{err}(t) = \log(\mathbf{R}_{err}(t)), \tag{12}$$

where $\mathbf{r}_{err}(t)$ and $\mathbf{R}_{err}(t)$ are the $\mathbb{R}^3$ and $\mathrm{SO}(3)$ elements of $\mathbf{T}_{err}(t)$. The set of parameters, $\mathbf{x}$, that we wish to estimate are the control points of the split representation of $\mathbf{T}_{wr}(t)$ and the extrinsic calibration parameters in $\mathbf{T}_{cr}$,

$$\mathbf{x} = \{\mathbf{p}_0, \quad \ldots, \quad \mathbf{p}_N, \quad \mathbf{R}_0, \quad \ldots, \quad \mathbf{R}_N, \quad \mathbf{R}_{cr}, \quad \mathbf{r}_c^{rc}\}. \tag{13}$$

Our optimization problem is then to find $\mathbf{x}^*$ that minimizes the following cost function:

$$\begin{aligned}\mathcal{J}(\mathbf{x}) = &\sum_{i=1}^{l} \mathbf{e}_v^T(t_i)\boldsymbol{\Sigma}_v^{-1}(t_i)\mathbf{e}_v(t_i)\\ &+ \sum_{j=1}^{m} \mathbf{e}_p^T(t_j)\boldsymbol{\Sigma}_p^{-1}(t_j)\mathbf{e}_p(t_j),\end{aligned} \tag{14}$$

where l and m are, respectively, the number of radar velocity measurements and camera pose measurements.

E. Implementation Details

Our approach to estimate the velocity of the radar unit involves finding the velocity vector that best fits a series of measured range-rate vectors. To do so, we use an algorithm and software package developed by Stahoviak et al. called 'Goggles' [3].[1] The Goggles algorithm applies MLESAC to find an inlier set of radar velocity measurements. The final velocity estimate is calculated using orthogonal distance regression on this inlier set of velocities.

We solve the full batch nonlinear optimization problem to determine the extrinsic parameters using the Levenberg-Marquardt implementation available in the Ceres solver [19]. Ceres' auto-differentiation capability is applied to calculate the Jacobians of the error equations. To manipulate the B-splines, we rely on the library from Sommer et al. [16].[2] Our translation and rotation splines have a spline order of $k = 4$.

IV. Observability Analysis

In order to estimate the calibration parameters, the system must be observable (or, equivalently for our batch formulation, identifiable). In Section IV-A, we make use of the observability rank condition criterion defined by Hermann and Krener [20] to prove that the calibration and scale estimation problem is observable. It is well known that, in the absence of metric distance information, absolute scale cannot be recovered from monocular camera measurements alone [21]. We show below that, given radar velocity data, it is possible to identify both the calibration parameters and the visual scale factor *without* knowledge of the (metric) distances between visual landmarks. It follows that radar-to-camera calibration, in the general case, does not require a specialized camera calibration target (or any other external source of scale information). We are concerned with the following set of parameters:

$$\mathbf{x} = \{\mathbf{r}_c^{rc}, \quad \mathbf{R}_{cr}, \quad \alpha\}, \tag{15}$$

where α is the unknown scale factor that appears in the camera pose measurement. A brief degeneracy analysis of the calibration problem, which identifies conditions that result in a loss of observability, is provided in Section IV-B.

A. Observability of Radar-to-Camera Extrinsic Calibration

We follow an approach similar to that in [22] and note that the (scaled) linear and angular velocities of the camera can be determined by taking the time derivatives of the camera pose measurements. Also, Stahoviak has shown that the 3D velocity of the radar (in the radar frame) can be recovered from three non-coplanar range-rate measurements [3]. These quantities can be related through rigid-body kinematics,

$$\mathbf{h}_i = \alpha\mathbf{v}_c^{cw} = \alpha(\mathbf{R}_{cr}\mathbf{v}_r^{rw} - \boldsymbol{\omega}_c^{\wedge}\mathbf{r}_c^{rc}), \tag{16}$$

where $\mathbf{h}_i$ is the scaled linear velocity of the camera and $\boldsymbol{\omega}_c$ is the angular velocity of the camera, both relative to the camera frame. To decrease the notational burden going forward, we drop the superscripts and subscripts defining the velocities and extrinsic transform parameters. The gradient of the zeroth-order Lie derivative of the ith measurement is

$$\nabla_{\mathbf{x}} L_0\mathbf{h}_i = \begin{bmatrix}-\alpha\boldsymbol{\omega}_i^{\wedge} & -\alpha(\mathbf{R}\mathbf{v}_i)^{\wedge}\mathbf{J} & \mathbf{R}\mathbf{v}_i - \boldsymbol{\omega}_i^{\wedge}\mathbf{r}\end{bmatrix}, \tag{17}$$

where $\mathbf{J}$ is the Lie algebra left Jacobian of $\mathbf{R}_{cr}$ [18]. Since the parameters of interest are constant with respect to time, we are able to stack the gradients of several Lie derivatives (at different points times) to form the observability matrix,

$$\mathbf{O} = \begin{bmatrix}\nabla_{\mathbf{x}} L_0\mathbf{h}_1\\ \nabla_{\mathbf{x}} L_0\mathbf{h}_2\\ \nabla_{\mathbf{x}} L_0\mathbf{h}_3\end{bmatrix}, \tag{18}$$

which has full column rank when three or more sets of measurements are available (we omit the full proof for brevity). We note that the analysis is simplified by considering the measurement equation only, and at different points in time. However, it is also possible to show that the system is instantaneously locally weakly observable when the sensor platform undergoes both linear and angular accelerations (again, we omit this proof due to space).

B. Degeneracy Analysis

The conditions under which a loss of observability (identifiability) may occur can be determined by examining the nullspace of the observability matrix. In this section, we consider the scale parameter to be known, which removes the last column of the matrix defined by Eq. 17—in turn,

[1] Available at https://github.com/cstahoviak/goggles

[2] Available at https://gitlab.com/VladyslavUsenko/basalt-headers.git

only two sets of measurements are required. The nullspace of $\nabla_{\mathbf{x}} L_0 \mathbf{h}_i$ contains the vectors

$$\mathbf{U}_i = \begin{bmatrix} \boldsymbol{\omega}_i & \mathbf{0} & (\mathbf{I} - \frac{\boldsymbol{\omega}_i \boldsymbol{\omega}_i^T}{\|\boldsymbol{\omega}_i\|^2})\mathbf{R}\mathbf{v}_i \\ \mathbf{0} & \mathbf{J}^{-1}\mathbf{R}\mathbf{v}_i & (\mathbf{I} - \frac{\mathbf{J}^{-1}\mathbf{R}\mathbf{v}_i (\mathbf{J}^{-1}\mathbf{R}\mathbf{v}_i)^T}{\left\|\mathbf{J}^{-1}\mathbf{R}\mathbf{v}_i\right\|^2})\mathbf{J}^{-1}\boldsymbol{\omega}_i \end{bmatrix}, \tag{19}$$

where each column of $\mathbf{U}_i$ defines one null vector. To ensure that the stacked observability matrix formed from $\nabla_{\mathbf{x}} L_0 \mathbf{h}_1$ and $\nabla_{\mathbf{x}} L_0 \mathbf{h}_2$ has full column rank (i.e., that the nullspace contains the zero vector only), the following constraints must be satisfied, at minimum:

$$\begin{aligned} \boldsymbol{\omega}_2 \times \boldsymbol{\omega}_1 &\neq 0, \\ \mathbf{v}_2 \times \mathbf{v}_1 &\neq 0. \end{aligned} \tag{20}$$

The constraints defined by Eq. 20 show that the system must rotate about and translate along two non-collinear axes at different points in time. The rotation constraint is expected because our problem is similar to the one defined by Brookshire and Teller in [23]. However, the angular velocity of the radar unit cannot be measured directly, which leads to the second excitation requirement. Additional constraints can be generated from the third column of Eq. 19, but these motions are more difficult to characterize; we posit, based on our experiments, that these constraints are less likely to be violated in practice.

V. Experiments and Results

In general, our algorithm can be applied to any 3D radar and egomotion sensor pair, but our experimental focus is on 3D radar-to-monocular camera extrinsic calibration. For convenience, in this work, we estimate the camera pose relative to a 12×10 planar checkerboard calibration target of known size. However, as shown in Section IV, knowledge of metric scale is not required—the camera must simply view a sufficient number of features (three or more) that lie in a general configuration in the environment.

Below, we present a series of synthetic and real world calibration experiments to evaluate the performance of our algorithm. In Section V-A, we empirically analyze the sensitivity of the algorithm to measurement noise when applied to synthetic data. In Section V-B, we demonstrate that our approach improves upon hand-measured calibration and compares favourably with the algorithm of Peršić et al. [24], although our approach does not require specialized radar retroreflectors.

A. Synthetic Data

Our simulation environment is shown in Fig. 2. In order to ensure sufficient excitation of the system, the sensor platform trajectory has non-zero linear and angular acceleration about all three axes in the radar sensor frame; see the bottom of Fig. 2. We added zero-mean Gaussian noise to each radar and camera measurement, with magnitudes similar to the noise levels identified in our real-world experiments.

Simulation results show that our algorithm is accurate in the low-noise regime, but that the performance degrades as the amount of noise in the radar velocity measurements increases (see Figure 3). We found that the average standard deviations of our real-world radar velocity estimates were 0.03, 0.06, and 0.1 m/s in the x, y, and z directions, respectively. As a result, our noisiest simulation experiment represents a worst-case calibration scenario, because the experiment uses twice the amount of noise as found in our true radar velocity data. Overall, the proposed calibration algorithm shows robustness to significant noise—we are able to successfully calibrate in all of our trials despite very large worst-case noise levels.

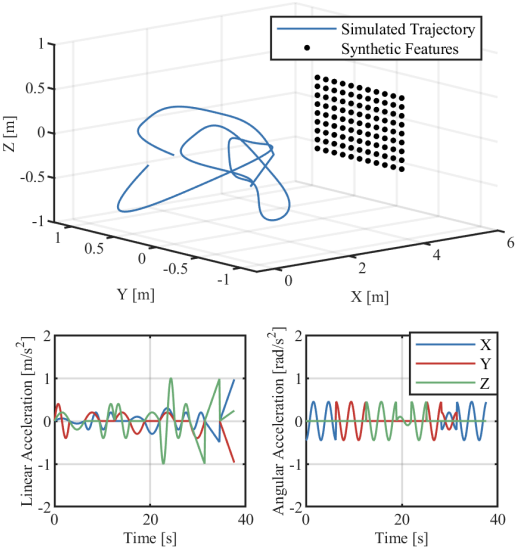

Fig. 2. Experimental setup for our simulation studies. The calibration rig rotates while moving along the blue trajectory. The black dots represent the internal corners of a 12-by-10 checkerboard with squares that are 9.9 cm by 9.9 cm in size, the same as those of our physical checkerboard.

B. Real-World Experiments

We collected a real-world dataset that allowed us to compare the performance of our algorithm to the 3D reprojection-based algorithm of Peršić et al. [24]. Our data collection rig (shown in Figure 4) carried: (i) a PointGrey BFLY-U3-23S6M-C global shutter camera with a Kowa C-Mount 6 mm fixed-focus lens ($96.8° \times 79.4°$ field of view) and (ii) a Texas Instruments AWR1843BOOST 3D radar unit. Both sensors operated at approximately 10 Hz. Data were captured and stored by an on-board Raspberry Pi 4 Model B. The camera intrinsic and lens distortion parameters were obtained using the Kalibr toolbox [25] prior to conducting the experiments. We performed a rough, ad hoc temporal alignment of the radar and camera data before running our optimization algorithm. Additionally, the extrinsic calibration (translation and rotation) parameters were carefully measured by hand for comparison.

Experiments were conducted outdoors to mitigate (to some extent) radar multipath reflections and other detrimental effects. We placed five specialized hybrid radar-camera targets [11] in the environment for validation purposes and for comparison with the calibration method in [24]. However, we emphasize that our algorithm does not specifically make

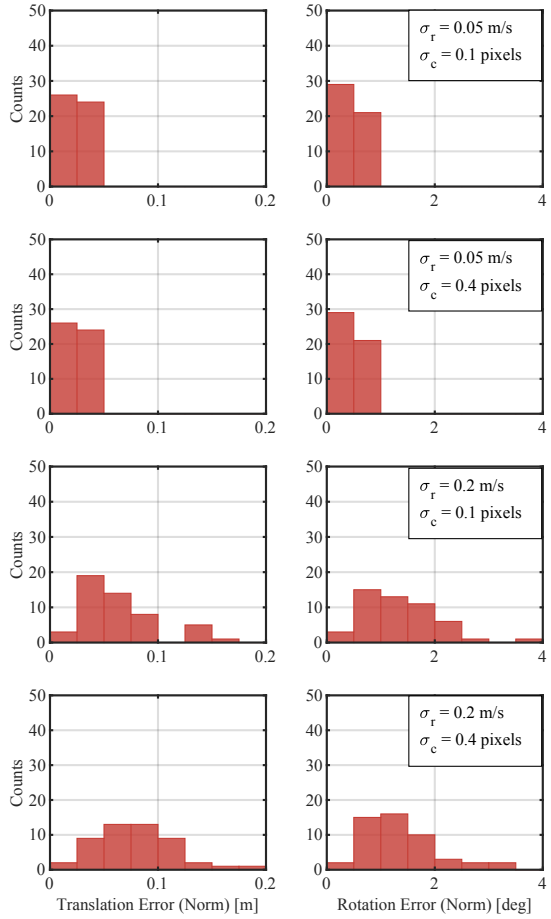

Fig. 3. Left: histograms of translation error norm between estimated and ground truth calibration parameters for different amounts of simulated radar velocity and image pixel noise. Right: histograms of rotation errors. The rotation error is the magnitude of the angle that aligns the estimated and true radar frames. For each noise combination, 50 test cases were run.

use of the retroreflective radar targets; the velocity of the radar can be determined independently.

We evaluated the performance of the calibration algorithm by measuring target reprojection error. We placed an AprilTag [26] on each radar-camera target in the environment, enabling us to estimate the 3D positions of the targets. Using the extrinsic transform obtained via a given calibration method, the radar measurement of the target can be projected into the camera reference frame. The distance between the observed 3D position of the target (from image data) and the projected radar estimate of the target position is the target reprojection error. Figure 5 shows the radar-to-camera reprojection error determined using three different calibration methods: hand-measurement, the 3D reprojection-based method of Peršić et al. [24], and our proposed method. Since the transform estimated by the 3D reprojection method in [24] optimally aligns the AprilTag positions with the projected radar measurements of the targets, this approach outperforms our algorithm according to this metric, as expected. However, the difference in the median reprojection error between our proposed method and that in [24] is less than 4 mm. In contrast to [24], our algorithm does not require any specialized radar targets in the general case.

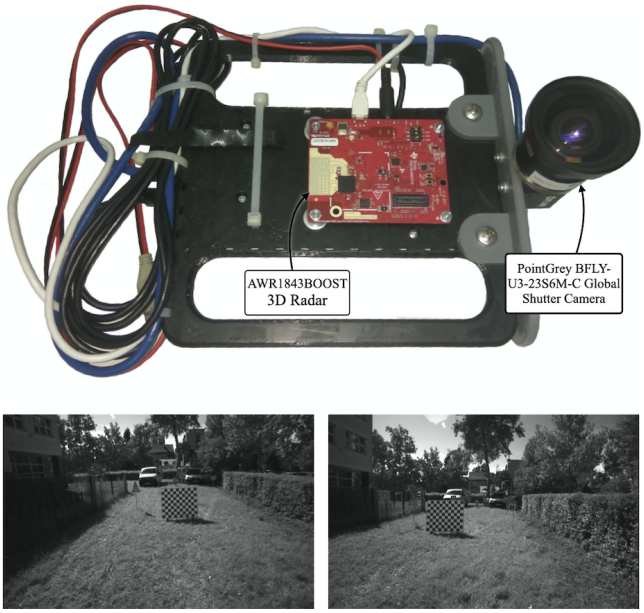

Fig. 4. The top image is a picture of the handheld data collection rig. The bottom two images show different perspectives of our data collection environment.

VI. Conclusion

In this paper, we described a novel continuous-time 3D millimetre-wavelength radar-to-camera extrinsic calibration algorithm. We showed that the problem is observable and derived the necessary conditions for calibration from radar velocity and camera pose measurements only. On synthetic data, our algorithm was shown to be accurate and reliable, but our sensitivity analysis indicated that performance depends on the amount of noise in the radar velocity measurements. Using data from a handheld sensor rig, we demonstrated that we are able to calibrate the extrinsic transform with an accuracy comparable to the method in [24] but without the need for retroreflectors. One future research direction is to investigate alternative cost functions that explicitly consider alignment errors (similar to [24]). Finally, joint spatiotemporal calibration [27] and monocular camera trajectory scale estimation, similar to [28], would be valuable extensions to our algorithm.

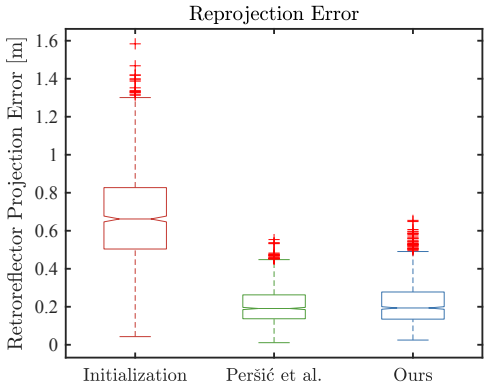

Fig. 5. The target reprojection error is shown for the following calibration methods: hand-measured, Peršić et al. [24], and our proposed method. All algorithms used the same dataset and all calibration results were obtained from a held-out dataset.

REFERENCES

[1] R. Gourova, O. Krasnov, and A. Yarovoy, "Analysis of rain clutter detections in commercial 77 GHz automotive radar," in *2017 European Radar Conference (EURAD)*, 2017, pp. 25–28.

[2] M. A. Richards, J. A. Scheer, and W. A. Holm, Eds., *Principles of Modern Radar: Basic principles*, ser. Radar, Sonar & Navigation. Institution of Engineering and Technology, 2010.

[3] C. C. Stahoviak, "An instantaneous 3D ego-velocity measurement algorithm for frequency modulated continuous wave (FMCW) doppler radar data," Master's thesis, University of Colorado at Boulder, 2019.

[4] S. Sugimoto, H. Tateda, H. Takahashi, and M. Okutomi, "Obstacle detection using millimeter-wave radar and its visualization on image sequence," in *International Conference on Pattern Recognition (ICPR)*, 2004, pp. 342–345.

[5] 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.

[6] D. Y. Kim and M. Jeon, "Data fusion of radar and image measurements for multi-object tracking via Kalman filtering," *Information Sciences*, vol. 278, pp. 641–652, 2014.

[7] J. Kim, D. S. Han, and B. Senouci, "Radar and vision sensor fusion for object detection in autonomous vehicle surroundings," in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, 2018, pp. 76–78.

[8] T. Kim, S. Kim, E. Lee, and M. Park, "Comparative analysis of RADAR-IR sensor fusion methods for object detection," in *2017 17th International Conference on Control, Automation and Systems (ICCAS)*, 2017, pp. 1576–1580.

[9] G. El Natour, O. Ait Aider, R. Rouveure, F. Berry, and P. Faure, "Radar and vision sensors calibration for outdoor 3D reconstruction," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 2084–2089.

[10] J. Domhof, J. F. P. Kooij, and D. M. Gavrila, "An extrinsic calibration tool for radar, camera and lidar," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 8107–8113.

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

[12] J. Oh, K. Kim, M. Park, and S. Kim, "A comparative study on camera-radar calibration methods," in *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2018, pp. 1057–1062.

[13] C. Schöller, M. Schnettler, A. Krämmer, G. Hinz, M. Bakovic, M. Güzet, and A. Knoll, "Targetless rotational auto-calibration of radar and camera for intelligent transportation systems," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 3934–3941.

[14] J. Peršić, L. Petrović, I. Marković, and I. Petrović, "Online multi-sensor calibration based on moving object tracking," *Advanced Robotics*, vol. 35, no. 3-4, pp. 130–140, 2021.

[15] D. Kellner, M. Barjenbruch, K. Dietmayer, J. Klappstein, and J. Dickmann, "Joint radar alignment and odometry calibration," in *2015 18th International Conference on Information Fusion (Fusion)*, 2015, pp. 366–374.

[16] C. Sommer, V. Usenko, D. Schubert, N. Demmel, and D. Cremers, "Efficient derivative computation for cumulative b-splines on Lie groups," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11 145–11 153.

[17] P. Furgale, C. H. Tong, T. D. Barfoot, and G. Sibley, "Continuous-time batch trajectory estimation using temporal basis functions," *The International Journal of Robotics Research*, vol. 34, no. 14, pp. 1688–1710, 2015.

[18] T. D. Barfoot, *State estimation for robotics*. Cambridge University Press, 2017.

[19] S. Agarwal, K. Mierle, and Others, "Ceres solver," http://ceres-solver.org.

[20] R. Hermann and A. Krener, "Nonlinear controllability and observability," *IEEE Transactions on Automatic Control (TAC)*, vol. 22, no. 5, pp. 728–740, 1977.

[21] A. Chiuso, P. Favaro, Hailin Jin, and S. Soatto, "Structure from motion causally integrated over time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 523–535, 2002.

[22] M. Li and A. I. Mourikis, "Online temporal calibration for camera-imu systems: Theory and algorithms," *International Journal of Robotics Research*, vol. 33, no. 7, pp. 947–964, 2014.

[23] J. Brookshire and S. Teller, "Extrinsic calibration from per-sensor egomotion," *Robotics: Science and Systems VIII*, pp. 504–512, 2013.

[24] J. Peršić, L. Petrović, I. Marković, and I. Petrović, "Spatio-temporal multisensor calibration based on gaussian processes moving object tracking," To appear in: *IEEE Transactions on Robotics (TRO)*.

[25] J. Maye, P. Furgale, and R. Siegwart, "Self-supervised calibration for robotic systems," in *2013 IEEE Intelligent Vehicles Symposium (IV)*, 2013, pp. 473–480.

[26] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 3400–3407.

[27] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1280–1286.

[28] E. Wise, M. Giamou, S. Khoubyarian, A. Grover, and J. Kelly, "Certifiably optimal monocular hand-eye calibration," in *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2020, pp. 271–278.