Multi-State Tightly-Coupled EKF-Based Radar-Inertial Odometry With Persistent Landmarks

Jan Michalczyk[1], Roland Jung, Christian Brommer and Stephan Weiss

***Abstract*— In this paper, we present a Radar-Inertial Odometry (RIO) approach that utilizes performance improving modules, enhanced for the sparse and noisy radar signals, from the vision community in order to estimate the full 6DoF pose and 3D velocity of a robot in an unprepared environment. Our method leverages a multi-state approach in which we make use of several past robot poses and trails of measurements from a lightweight and inexpensive Frequency Modulated Continuous Wave (FMCW) radar sensor. Furthermore, in our estimation framework we include a method for promoting measurement trails to *persistent landmarks* which correspond to salient features in the environment. In an Extended Kalman Filter (EKF) framework, we fuse the range measurements to the persistent landmarks, trails, and the velocity measurements of the detected 3D points together with the Inertial Measurement Unit (IMU) readings. Our method is particularly relevant for (but not limited to) Unmanned Aerial Vehicles (UAV), enabling them to localize while performing missions in Global Navigation Satellite System (GNSS)-denied environments and, thanks to the properties of the radar sensor, in environments generally challenging for robot perception due to external factors such as smoke or extreme illumination. We show in real flight experiments the effectiveness of our estimator and compare it to the state-of-the-art.**

I. Introduction And Related Work

Accurate odometry estimation is a prerequisite for the autonomous operation of a robot in GNSS-denied and unprepared environments. Unmanned (aerial) vehicles can perform autonomous missions in such environments while possibly dealing with the external factors rendering their perception of those environments even more challenging. Such factors may include extreme lighting conditions or presence of aerosols in the air such as fog or smoke. In such scenarios the robustness of the used sensors is of paramount importance.

In the light of environmental resilience requirements, a sensor suite composed of an FMCW radar and an IMU raises an increasing interest in the UAV research community to form the backbone of the localization system.

FMCW radar does not suffer from the above-mentioned environment-induced issues [1] yet it is small in size and deliver measurements at higher or similar frequencies as other sensors commonly used for navigation. Radars report distance, Doppler velocity, and to a limited extent angular measurements of reflecting points in the environment. They have been thrivingly used in automotive industry [2], [3], [4]. Millimeter-wave technology triggered their miniaturization and increased their accuracy which prompted their application on UAVs as sensors used for correcting IMU drift acquired during integration.

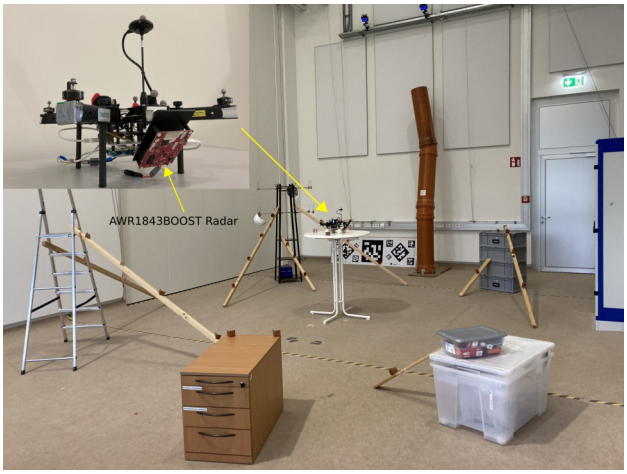

Fig. 1. Experimental platform used in this work with the FMCW radar sensor and the indoor space where the experiments were conducted. Note the mounting of the sensor tilted at 45° angle. Reflective clutter was scattered randomly on the scene. No global position (nor attitude) information of any sort about scattered objects was used in our approach.

In [5] the authors present a nonlinear optimization based approach in which a cost function is minimized over a moving window of past IMU and FMCW radar Doppler velocity measurements to estimate the ego-velocity of an UAV. The demonstrated method is shown to be effective in conditions challenging for Visual-Inertial Odometry (VIO), nevertheless, it does not allow for precise full 6DoF pose estimation. Authors in [6] and [7] describe an approach, where no data association between consecutive radar pointclouds is needed, only Doppler velocities of points in the current scan are used. Significant drift in the estimates computed with these methods is countered by using a barometer sensor. In [8], the two aforementioned methods are further developed by making use of the Manhattan world assumptions on the environment to limit the yaw angle drift. These assumptions render the demonstrated method to perform on-par with VIO. A number of other odometry methods are reported in which a high-resolution, mechanically rotating radar is used in automotive context. In [9] a deep learning based method is outlined, which enables the prediction of robust features in radar scans. The features are subsequently used to calculate the relative transform between the scans. Another work making use of deep learning for feature detection but this time combining it with the probabilistic trajectory estimation is shown in [10]. In [11] a featureless approach

[1] All authors are with the Control of Networked Systems Group, University of Klagenfurt, Austria {firstname.lastname}@ieee.org

This research received funding from the Austrian Ministry of Climate Action and Energy (BMK) under the grant agreement 880057 (CARNIVAL).

Pre-print version, accepted January/2023, DOI follows ASAP ©IEEE.

employing the Fourier-Mellin transform is used to estimate from an entire radar scan the robot translation and rotation. The authors of [12] propose an approach in which local geometrical relationships of detected consecutive radar pointclouds are leveraged in features association. In [13] the authors robustify the feature extraction scheme shown in [12] by including prior assumption on the robot motion to reject outliers. In both methods Singular Value Decomposition (SVD) is used to estimate the transform between the scan poses from the features. [9], [10], [11], [12], [13] make use of bulky and expensive mechanically rotating radar not readily applicable on a UAV. In [14] the authors adapt the scan matching method from [12] to a lightweight FMCW radar showing promising results when fused in loosely-coupled fashion with IMU, applied however, on a mobile platform slowly moving over a flat scene, which reduces the complexity of the scenario as compared to an UAV. In [15], the authors outline an effective EKF-based tightly-coupled RIO approach in which a single past robot pose is used to align two consecutive pointclouds, from which matched 3D points are used for update together with Doppler velocities.

In the present paper, we significantly reduced the final drift (by a factor of 4 in average) and increased the accuracy (in terms of Mean Absolute Error (MAE) by a factor of 2) of RIO compared to the single-frame approach described in [15] and demonstrate it in real flights. To this end, we employ the stochastic cloning [16] for augmenting the state with a chosen number of past robot poses and corresponding radar scans (3D pointclouds) in a First In First Out (FIFO) buffer from which measurement trails are constructed. Trails matched consistently over a given amount of time are promoted to persistent landmarks and added to the state vector. In addition to these distance measurements to landmarks and trails, we also use Doppler velocity of points from the current radar scan. We fuse all measurements in a tightly-coupled formulation in our EKF setup. The tight coupling permits the integration of single distance and velocity measurements during update steps. This property obviates any limitations on required minimal number of matches (as it is e.g., needed for a prior Iterative Closest Point (ICP) and subsequent loose coupling of the resulting delta-pose in the EKF). This is a particularly strong advantage in view of robustness and accuracy over loosely coupled approaches since, e.g., ICP [17] works poorly on noisy and sparse FMCW radar point clouds. Note that our RIO method makes no assumptions on the environment and makes use of no other sensors than IMU and a lightweight millimeter-wave FMCW radar providing sparse and noisy 3D point clouds along with Doppler velocities of the detected points. It is suitable for a UAV and real-time capable. Our main contributions are:

- Application of stochastic cloning for inclusion of *multiple past robot poses* used for formulating an update equation on the accurately measured *distances* to multiple points in radar measurement trails.
- Implementation of persistent radar landmarks for increased estimation accuracy reducing global pose drift by a factor 4 in average compared to state-of-the-art.
- Past radar feature matching and efficient trail generation/handling for using past measurement-trails.
- Comparison against current methods [15], [6] and [7] using similar setup in real-world flight experiments.

Note that the leveraged techniques from the vision community [18], the sparse and noisy radar 3D (versus 2D in vision) measurements require important enhancements. This includes a different definition and treatment of 3D instead of 2D trails, a measurement definition along the most precise dimension of the sensor (i.e., radial distance), and the inclusion of Doppler velocity measurements. On the other hand, the multi-state approach for radar inherently handles hovering situations where the 2D vision measurements require special treatment. This paper is organized as follows. Section II introduces the preliminaries of our work. In subsection II-A we introduce the notation used, the system overview in subsection II-B. In section III we describe our RIO method. Subsection III-A outlines how the state vector and covariance matrix of our system are augmented using stochastic cloning. Subsection III-B describes how measurement trails are used in the update step. Subsection III-C explains how persistent landmarks are implemented and used for update. We summarize our estimator in subsection III-D. Experiments conducted in order to validate the proposed method are outlined in Section IV. In subsection IV-A we report the experimental setup used during the experiments and the subsection IV-B presents the results of the evaluation. Finally, we present conclusions in Section V.

II. Preliminaries

A. Notation

We use the generally known notation for the easier following of the paper. A normally distributed multivariate variable is defined as $\mathbf{X}_i \sim \mathcal{N}(\mathbf{x}_i, \mathbf{\Sigma}_{ii})$, with a mean $\mathbf{x}_i$ and covariance (uncertainty) $\mathbf{\Sigma}_{ii}$, which is called the *belief* of i. Due to the symmetry property of covariances, we abbreviate lower triangular elements by a $\{\bullet\}$. Names of reference frames are capitalized and calligraphic, e.g., $\{\mathcal{I}\}$ for IMU. A pose between the reference frames $\mathcal{A}$ and $\mathcal{B}$ is defined as ${}^{\mathcal{A}}\mathbf{T}_{\mathcal{B}} = \begin{bmatrix} {}^{\mathcal{A}}\mathbf{R}_{\mathcal{B}} & {}^{\mathcal{A}}\mathbf{p}_{\mathcal{B}} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \in \mathrm{SE}(3)$, with $\mathbf{R} \in \mathrm{SO}(3)$ and $\mathbf{p} \in \mathbb{R}^3$. The transformation of a coordinate vector ${}^{\mathcal{C}}_{\mathcal{C}}\mathbf{p}_{P_1}$ pointing from the origin of the reference frame $\mathcal{C}$ to a point P_1, expressed in $\mathcal{C}$, can be transformed into the frame $\mathcal{A}$ by $\begin{bmatrix} {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{p}_{P_1} \\ 1 \end{bmatrix} = {}^{\mathcal{A}}\mathbf{T}_{\mathcal{C}} \begin{bmatrix} {}^{\mathcal{C}}_{\mathcal{C}}\mathbf{p}_{P_1} \\ 1 \end{bmatrix}$ (read as ${}^{from}_{in}\mathbf{x}_{\,to}$). Rotations are stored as unit quaternion $\bar{\mathbf{q}} \in \mathrm{SO}(3)$ with $\|\bar{\mathbf{q}}\| = 1$ allowing a direct mapping between rotation matrices and unit Hamiltonian quaternions by ${}^{\mathcal{A}}\mathbf{R}_{\mathcal{B}} = \mathbf{R}\{{}^{\mathcal{A}}\bar{\mathbf{q}}_{\mathcal{B}}\} \in \mathrm{SO}^3$ and ${}^{\mathcal{A}}\bar{\mathbf{q}}_{\mathcal{B}} = \bar{\mathbf{q}}\{{}^{\mathcal{A}}\mathbf{R}_{\mathcal{B}}\}$ [19]. $\mathbf{I}$ is the identity matrix. The *a priori* and *a posteriori* of a belief are indicated by a $\{\bullet\}^{(-)}$ and $\{\bullet\}^{(+)}$, respectively. $\{\bullet\}^{\#}$ specifies measured (perturbed) quantities. For vectors and block matrices, semicolons and colons improve the readability such that $[\mathbf{A};\mathbf{B}] \equiv \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}$ and $[\mathbf{A},\mathbf{B}] \equiv \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix}$.

B. System Overview

Our RIO method is based on an error-state EKF formulation [20] which uses an IMU as the primary sensor for the state propagation. Updates are performed with the FMCW radar measurements, which consist of sparse and noisy 3D pointclouds and relative radial velocities of detected points. The principle of sensing of the radar we use is explained briefly in [15]. Every time a radar measurement is obtained, we augment the state of our EKF filter with the pose of the robot at which the measurement took place using stochastic cloning as described in Section III-A. New poses are appended to the buffer of past poses in a FIFO fashion. The maximum number of cloned poses is defined by the parameter N. From taken measurements, we construct and maintain a set of trails which record the continuous detectability of 3D points by the radar sensor, which have maximum trail length N. Meaning that, every matched point keeps a history of its detected positions and every element of this history refers to a cloned robot pose at which the detection was taken. Such a point with a history of detections is referred to as a trail and it is kept in memory as long as it is actively matched to a point in the current scan following our sparse radar pointcloud matching method described in [15]. If it is not matched, it is inactive and thus removed. For 3D point matching, we use the latest detection of each trail along with the robot pose at which it was taken, together with the pose at which the current scan was taken. Specifically, we use these two poses to spatially align the current radar scan with the trails and trigger the matching procedure on such aligned matches. Once matched, the whole trail history is used to form the residual vector in the EKF update. Next, we use projections of the current robot velocity onto normal vectors to all points detected in the current radar scan together with their measured velocities to further augment the residual vector. The final component of the residual vector comes from using persistent landmarks. Namely, in the case a trail has been continuously seen for N times, it is removed from the set of trails, added to the state vector as persistent landmark and matched to the detections. Residual vectors are then used in the update step to estimate the mean of the error-state, which is injected into the regular state. The coordinate frames arrangement for measurements in our system is shown in Fig. 3.

III. Multi-State Radar-Inertial State Estimation With Persistent Landmarks

The state vector $\mathbf{x}$ in our filter is defined as follows:

$$\mathbf{x} = [\mathbf{x}_I; \mathbf{x}_C; \mathbf{x}_L] = \Big[\big[{}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}}; {}^{\mathcal{G}}\bar{\mathbf{q}}_{\mathcal{I}}; {}^{\mathcal{G}}\mathbf{v}_{\mathcal{I}}; \mathbf{b_a}; \mathbf{b}_{\boldsymbol{\omega}}\big] ; \big[{}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}_1}; {}^{\mathcal{G}}\bar{\mathbf{q}}_{\mathcal{I}_1}; \dots; {}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}_N}; {}^{\mathcal{G}}\bar{\mathbf{q}}_{\mathcal{I}_N}\big] ; \big[{}^{\mathcal{G}}\mathbf{p}_{\mathcal{L}_1}; \dots; {}^{\mathcal{G}}\mathbf{p}_{\mathcal{L}_M}\big] \Big] \tag{1}$$

with the IMU state $\mathbf{x}_I$, the stochastically cloned states $\mathbf{x}_C$ of the IMU poses corresponding to the previous radar measurements as described in Section III-A and the set of persistent landmarks $\mathbf{x}_L$ as described in Section III-C. The previous radar measurements (point cloud of reflecting objects and their Doppler velocities) are not part of the state vector. ${}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}}$, ${}^{\mathcal{G}}\mathbf{v}_{\mathcal{I}}$, and ${}^{\mathcal{G}}\bar{\mathbf{q}}_{\mathcal{I}}$ are the position, velocity, and orientation of the IMU/body frame $\{\mathcal{I}\}$ with respect to the navigation frame $\{\mathcal{G}\}$, respectively. $\mathbf{b}_{\boldsymbol{\omega}}$ and $\mathbf{b_a}$ are the measurement biases of the gyroscope and accelerometer, respectively. $\big[{}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}_n}, {}^{\mathcal{G}}\bar{\mathbf{q}}_{\mathcal{I}_n}\big]$ with $n = 1, \dots, N$ define a set of past IMU poses with respect to the navigation frame $\{\mathcal{G}\}$ at the moments of past radar measurements. ${}^{\mathcal{G}}\mathbf{p}_{\mathcal{L}_m}$ with $m = 1, \dots, M$ define the position of persistent landmarks $\{\mathcal{L}\}$ with respect to the navigation frame $\{\mathcal{G}\}$. We use $\big[{}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}_1}, {}^{\mathcal{G}}\bar{\mathbf{q}}_{\mathcal{I}_1}\big]$ (corresponding to the newest coordinates of the trails) for ad-hoc point correspondence generation such that we do not need to keep 3D points in the state vector in order to use distance based measurements (Section III-A).

The evolution of the state is expressed by the following differential equations:

$$\begin{aligned}
{}^{\mathcal{G}}\dot{\mathbf{p}}_{\mathcal{I}} &= {}^{\mathcal{G}}\mathbf{v}_{\mathcal{I}}, \\
{}^{\mathcal{G}}\dot{\mathbf{v}}_{\mathcal{I}} &= {}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}} \left({}_{\mathcal{I}}\mathbf{a}^{\#} - \mathbf{b_a} - \mathbf{n_a}\right) + {}_{\mathcal{G}}\mathbf{g}, \\
{}^{\mathcal{G}}\dot{\mathbf{R}}_{\mathcal{I}} &= {}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}} \left[{}_{\mathcal{I}}\boldsymbol{\omega}^{\#} - \mathbf{b}_{\boldsymbol{\omega}} - \mathbf{n}_{\boldsymbol{\omega}}\right]_{\times}, \\
\dot{\mathbf{b}}_{\mathbf{a}} &= \mathbf{n}_{\mathbf{b_a}}, \dot{\mathbf{b}}_{\boldsymbol{\omega}} = \mathbf{n}_{\mathbf{b}_{\boldsymbol{\omega}}}, {}^{\mathcal{G}}\dot{\mathbf{p}}_{\mathcal{I}_n} = \mathbf{0}, {}^{\mathcal{G}}\dot{\mathbf{R}}_{\mathcal{I}_n} = \mathbf{0}, \\
{}^{\mathcal{G}}\dot{\mathbf{p}}_{\mathcal{L}_m} &= \mathbf{0}
\end{aligned} \tag{2}$$

where $n = 1, .., N$ refers to the most recent past IMU poses which are not changing in time, $m = 1, .., M$ refer to M most recent estimated positions of landmarks, ${}_{\mathcal{I}}\mathbf{a}^{\#}$

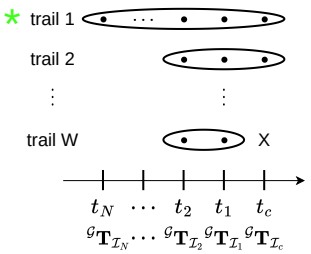

Fig. 2. Measurement trails used to record a history of detections. The green star marks a trail that was classified as a persistent landmark. X means that the trail W has not been matched to any point in the current scan at t_c and will thus be deleted. Dot in each trail is a single past detection. As shown below time instant t, points in the trail have associated cloned robot poses at which the measurements were taken.

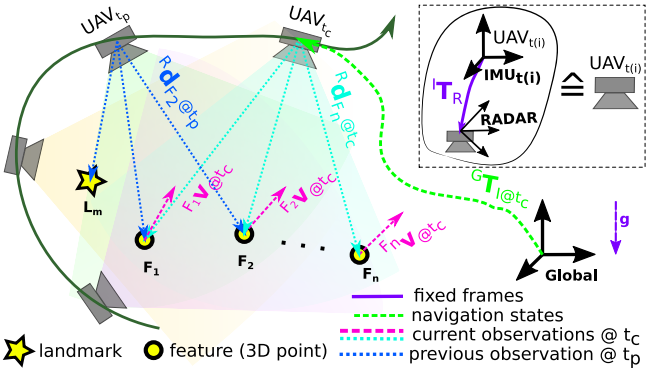

Fig. 3. Multiple consecutive past UAV poses are used in the distance measurements models employing persistent landmarks and trails of measurements. In the velocity measurement model, only the readings from the current pose are used.

and ${}_{\mathcal{I}}\boldsymbol{\omega}^{\#}$ are the accelerometer and gyroscope measurements of the IMU with a white measurement noise $\mathbf{n_a}$ and $\mathbf{n}_{\boldsymbol{\omega}}$. $\mathbf{n_{b_a}}$ and $\mathbf{n_{b_\omega}}$ are assumed to be white Gaussian noise to model the bias change over time as a random process. The gravity vector is assumed to be aligned with the z-axis of the navigation frame ${}_{\mathcal{G}}\mathbf{g} = [0, 0, 9.81]^{\mathsf{T}}$.

Since we use an error-state EKF formulation we introduce the error state vector from the states defined in Equation (1):

$$\tilde{\mathbf{x}}^{\mathsf{T}} = [\tilde{\mathbf{x}}_I; \tilde{\mathbf{x}}_C; \tilde{\mathbf{x}}_L] = \Big[\Big[{}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{I}}; {}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{I}}; {}^{\mathcal{G}}\tilde{\mathbf{v}}_{\mathcal{I}}; \tilde{\mathbf{b}}_{\mathbf{a}}; \tilde{\mathbf{b}}_{\boldsymbol{\omega}} \Big]; \Big[{}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{I}_1}; {}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{I}_1}; \ldots; {}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{I}_N}; {}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{I}_N} \Big]; \big[{}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{L}_1}; ...; {}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{L}_M} \big] \Big] \tag{3}$$

For translational components, e.g., the position, the error is defined as ${}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{I}} = {}^{\mathcal{G}}\hat{\mathbf{p}}_{\mathcal{I}} - {}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}}$, while for rotations/quaternions it is defined as $\tilde{\bar{\mathbf{q}}} = \hat{\bar{\mathbf{q}}}^{-1} \otimes \bar{\mathbf{q}} = \big[1; \frac{1}{2}\tilde{\boldsymbol{\theta}}\big]$, with $\otimes$ and $\tilde{\boldsymbol{\theta}}$ being quaternion product and small angle approximation, respectively.

A. State Augmentation

In order to process relative measurements relating to estimates at different time instances, Roumeliotis and Burdick introduce the concept of Stochastic Cloning (SC) in [16]. To appropriately consider the correlations/interdependencies between the estimates from different time instances, an identical copy of the required states and their uncertainties is used to augment the state vector and the corresponding error-state covariance matrix. Given the error-state definition in Equation (3), $\tilde{\mathbf{x}}_C$ is defined as the error state of the stochastic clone of the IMU pose $\big[{}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}}; {}^{\mathcal{G}}\mathbf{q}_{\mathcal{I}}\big]$. As the newest cloned state is fully correlated with the IMU pose and remaining cloned states are correlated with each other, it leads to the following augmented covariance matrix of the corresponding error-state:

$$\tilde{\mathbf{x}} = [\tilde{\mathbf{x}}_I; \tilde{\mathbf{x}}_{C_1}; \ldots; \tilde{\mathbf{x}}_{C_n}; \tilde{\mathbf{x}}_L], \tag{4}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_I & \boldsymbol{\Sigma}_{IC_1} & \cdots & \boldsymbol{\Sigma}_{IC_n} & \boldsymbol{\Sigma}_{IL} \\ \bullet & \boldsymbol{\Sigma}_{C_1} & \cdots & \boldsymbol{\Sigma}_{C_1C_n} & \boldsymbol{\Sigma}_{C_1L} \\ \bullet & \bullet & \ddots & \vdots & \vdots \\ \bullet & \bullet & \bullet & \boldsymbol{\Sigma}_{C_n} & \boldsymbol{\Sigma}_{C_nL} \\ \bullet & \bullet & \bullet & \bullet & \boldsymbol{\Sigma}_L \end{bmatrix} \tag{5}$$

with $\boldsymbol{\Sigma}_I$ being the 15×15 uncertainty of the IMU error-state $\tilde{\mathbf{x}}_I$. $\boldsymbol{\Sigma}_L$ is the $3M \times 3M$ uncertainty of a set of M landmark error-states $\tilde{\mathbf{x}}_L = [\tilde{\mathbf{x}}_{L_1}; \ldots; \tilde{\mathbf{x}}_{L_m}]$. $\boldsymbol{\Sigma}_{C_1} = \boldsymbol{\Sigma}_{I_{\{\tilde{\mathbf{p}}, \tilde{\boldsymbol{\theta}}\}}}$ is the 6×6 uncertainty of the newly cloned IMU pose error state (which is fully correlated, thus $\boldsymbol{\Sigma}_{IC_1} = \boldsymbol{\Sigma}_{II_{\{\tilde{\mathbf{p}}, \tilde{\boldsymbol{\theta}}\}}}$). All cross-covariances of the current IMU pose are assigned to the cross-covariances of the newly cloned state: $\boldsymbol{\Sigma}_{C_1C_i} = \boldsymbol{\Sigma}_{I_{\{\tilde{\mathbf{p}}, \tilde{\boldsymbol{\theta}}\}}C_i}$ with $i = 2, \ldots, n$ and $\boldsymbol{\Sigma}_{C_1L} = \boldsymbol{\Sigma}_{I_{\{\tilde{\mathbf{p}}, \tilde{\boldsymbol{\theta}}\}}L}$. $\boldsymbol{\Sigma}_{C_n}$ is the 6×6 uncertainty of the oldest cloned IMU pose. $\boldsymbol{\Sigma}_{C_iC_j}$ is the cross-correlation between i-th and j-th cloned IMU pose, $\boldsymbol{\Sigma}_{IC_i}$ is the cross-correlation between the current IMU state and the i-th cloned IMU pose, and $\boldsymbol{\Sigma}_{C_iL}$ are the cross-correlations between the IMU poses and the landmarks.

The cloned poses and landmarks do not evolve with time, meaning no state transition (i.e., $\boldsymbol{\Phi}_{C_n}^{k+1|k} = \mathbf{I}$ with $n = 1, ..., N$ and $\boldsymbol{\Phi}_{L_m}^{k+1|k} = \mathbf{I}$ with $m = 1, ..., M$) and no process noise (i.e., $\mathbf{G}_C^{k+1|k} = \mathbf{0}$ and $\mathbf{G}_L^{k+1|k} = \mathbf{0}$) is applied, while the navigation states evolve with the IMU measurements. The linearized error state propagation can be derived as:

$$\begin{aligned} \tilde{\mathbf{x}}^{k+1} &= \boldsymbol{\Phi}^{k+1|k}\tilde{\mathbf{x}}^k + \mathbf{G}^{k+1|k}\mathbf{w}^k, \\ \begin{bmatrix} \tilde{\mathbf{x}}_I^{k+1} \\ \tilde{\mathbf{x}}_C^{k+1} \\ \tilde{\mathbf{x}}_L^{k+1} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\Phi}_I^{k+1|k} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_C^{k+1|k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi}_L^{k+1|k} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_I^k \\ \tilde{\mathbf{x}}_C^k \\ \tilde{\mathbf{x}}_L^k \end{bmatrix} \\ &\quad + \begin{bmatrix} \mathbf{G}_I^{k+1|k} \\ \mathbf{G}_C^{k+1|k} \\ \mathbf{G}_L^{k+1|k} \end{bmatrix} \mathbf{w}^k \\ &= \begin{bmatrix} \boldsymbol{\Phi}_I^{k+1|k} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_I^k \\ \tilde{\mathbf{x}}_C^k \\ \tilde{\mathbf{x}}_L^k \end{bmatrix} + \begin{bmatrix} \mathbf{G}_I^{k+1|k} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \mathbf{w}^k \end{aligned} \tag{6}$$

with the linearized state transition matrix $\boldsymbol{\Phi}$ and the linearized perturbation matrix $\mathbf{G}$ computed as explained by Weiss in [21] or related work. The full error-state uncertainty of Equation (5) can then be propagated as

$$\begin{aligned} \boldsymbol{\Sigma}^{k+1} &= \boldsymbol{\Phi}^{k+1|k}\boldsymbol{\Sigma}^k(\boldsymbol{\Phi}^{k+1|k})^{\mathsf{T}} + \mathbf{G}^{k+1|k}\mathbf{Q}^k(\mathbf{G}^{k+1|k})^{\mathsf{T}} \\ &= \begin{bmatrix} \boldsymbol{\Sigma}_I^{k+1} & \boldsymbol{\Phi}_I^{k+1|k}\boldsymbol{\Sigma}_{IC}^k\mathbf{I} & \boldsymbol{\Phi}_I^{k+1|k}\boldsymbol{\Sigma}_{IL}^k\mathbf{I} \\ \bullet & \boldsymbol{\Sigma}_C^k & \boldsymbol{\Sigma}_{CL}^k \\ \bullet & \bullet & \boldsymbol{\Sigma}_L^k \end{bmatrix} \end{aligned} \tag{7}$$

with $\mathbf{Q}$ being the discretized process noise matrix and $\boldsymbol{\Phi}_I^{k+1|k}$ the error-state transition matrix of the IMU error-state $\tilde{\mathbf{x}}_I$. This propagation allows us to rigorously reflect the cross-correlations between the landmark, the cloned states, and the evolved IMU states in our error-state formulation. The above described formalism enables us to correctly use the state variables in order to align the trails to the current scan prior to point matching as well as compute residuals during the update.

B. Multi-State Update With Measurement Trails

Given a set of matched 3D point-trails as in the Fig. 2, we now want to estimate the distances to the matched points in the current scan across all points contained in the trails history. For a single matched trail, using cloned poses in the buffer, we transform all points ${}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_j}^{t_p}$ from the trail history at time instance t_p, where $p = 1, \ldots, V$ and V is the length of the matched trail, to the current radar reference frame, considering the robot's spatial evolution:

$$\begin{aligned} {}^{\mathcal{R}}\mathbf{p}'^{t_p}_{\mathcal{P}_j} = {}^{\mathcal{I}}\mathbf{R}_{\mathcal{R}}^{\mathsf{T}} \Big(-{}^{\mathcal{I}}_{\mathcal{I}}\mathbf{p}_{\mathcal{R}} + ({}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}}^{t_c})^{\mathsf{T}} \Big(-{}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}}^{t_c} + \\ {}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}}^{t_p} + {}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}}^{t_p} \Big({}^{\mathcal{I}}_{\mathcal{I}}\mathbf{p}_{\mathcal{R}} + {}^{\mathcal{I}}\mathbf{R}_{\mathcal{R}}\, {}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_j}^{t_p} \Big) \Big) \Big) \end{aligned} \tag{8}$$

where ${}^{\mathcal{I}}\mathbf{R}_{\mathcal{R}}$ and ${}^{\mathcal{I}}\mathbf{p}_{\mathcal{R}}$ is the constant pose (orientation and position) of the radar frame with respect to the IMU frame. ${}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}}^{\{t_c, t_p\}}$ and ${}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}}^{\{t_c, t_p\}}$ are the IMU orientation and position corresponding to the trail history element at time t_p and current radar scan at t_c, with respect to the navigation frame $\{\mathcal{G}\}$. Similarly to [15], we transform the 3D point from Cartesian space to Spherical coordinates and only use

the most informative dimension, the distance for residual construction.

The estimated distance, which is compared to the current distance measurement, is calculated for each point in the trail history as the norm of the transformed point from t_p:

$$d_{\mathcal{P}_j} = \left\| {}^{\mathcal{R}}\mathbf{p}'^{t_p}_{\mathcal{P}_j} \right\| \tag{9}$$

where $d_{\mathcal{P}_j}$ is the distance to a single point in the matched trail history ${}^{\mathcal{R}}\mathbf{p}'^{t_p}_{\mathcal{P}_j}$ at t_p aligned to the current radar pose at t_c. Since this measurement relates to states from past time instances, stochastic cloning is necessary as introduced in Section III-A.

C. Update With Persistent Landmarks

When a trail has been continuously matched for a predefined amount of times in the past, it is promoted to a persistent landmark and added as such to the state vector. Specifically, after each update, the set of trails is scanned for elements which have been matched consecutively for N times. When a trail meets this criterion, it is used to initialize a persistent landmark in the state vector and the covariance matrix is augmented according to [22]. For convenience, we introduce $\mathbf{x}_D = [\mathbf{x}_I; \mathbf{x}_C]$. Blocks needed for augmenting the state vector and error-state covariance are shown in the Fig. 4 and are computed as:

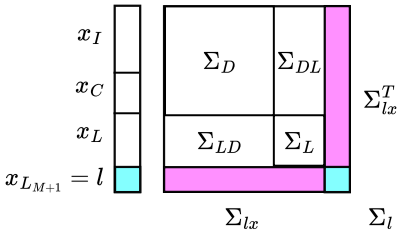

Fig. 4. Augmented nominal state and error-state covariance after adding a persistent landmark.

$$\mathbf{\Sigma}_l = \mathbf{H}_D \mathbf{\Sigma}_D \mathbf{H}_D^\mathsf{T} + \mathbf{H}_l \mathbf{R} \mathbf{H}_l^\mathsf{T}, \tag{10}$$

$$\mathbf{\Sigma}_{lx} = \mathbf{H}_D \mathbf{\Sigma}_{Dx} \tag{11}$$

with $\mathbf{\Sigma}_{Dx} = [\mathbf{\Sigma}_D, \mathbf{\Sigma}_{DL}]$, and $\mathbf{\Sigma}_{DL}$ being the cross-covariance between the IMU and IMU clones error-state vector segment, and the persistent landmarks. $\mathbf{R}$ is the covariance matrix of the measurement noise, $\mathbf{H}_D = \frac{\partial p}{\partial \tilde{\mathbf{x}}_D}$ and $\mathbf{H}_l = \frac{\partial p}{\partial \tilde{\mathbf{l}}}$ are the Jacobians of the inverse observation model of a 3D point radar measurement, p (Eq. 12), with respect to the IMU and IMU clones error-state variables $\tilde{\mathbf{x}}_D$, and the newly added landmark $\mathbf{l} = \mathbf{x}_{L_{M+1}}$, respectively. The inverse observation model of the 3D radar point in the navigation frame $\{\mathcal{G}\}$ is expressed as:

$${}^{\mathcal{G}}_{\mathcal{G}}\mathbf{p}_{\mathcal{L}_m} = p(\mathbf{x}, \mathbf{z}) = {}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}} \left({}^{\mathcal{I}}\mathbf{R}_{\mathcal{R}} {}^{\mathcal{R}}_{\mathcal{R}}\mathbf{p}_{\mathcal{L}_m} + {}^{\mathcal{I}}_{\mathcal{I}}\mathbf{p}_{\mathcal{R}} \right) + {}^{\mathcal{G}}_{\mathcal{G}}\mathbf{p}_{\mathcal{I}} \tag{12}$$

with ${}^{\mathcal{I}}\mathbf{R}_{\mathcal{R}}$ and ${}^{\mathcal{I}}_{\mathcal{I}}\mathbf{p}_{\mathcal{R}}$ being the pose between the IMU and radar sensor (which is assumed to be rigid and known a-priori), ${}^{\mathcal{R}}_{\mathcal{R}}\mathbf{p}_{\mathcal{L}_m}$ is the radar observation of the trail point in the current radar reference frame $\{\mathcal{R}\}$ with which an m-th landmark will be initialized, and ${}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}}$ and ${}^{\mathcal{G}}_{\mathcal{G}}\mathbf{p}_{\mathcal{I}}$ being the current pose of the IMU in the navigation frame. For readability, the estimate of the m-th landmark is abbreviated by $\mathbf{l}_m = {}^{\mathcal{G}}_{\mathcal{G}}\mathbf{p}_{\mathcal{L}_m}$.

When a persistent landmark does not have a match within the current radar scan, then it is discarded from the state vector and the covariance matrix is shrunk accordingly.

Finally, the estimated distance used for the update is computed according to:

$$\mathbf{l}'_m = {}^{\mathcal{R}}_{\mathcal{R}}\mathbf{p}_{\mathcal{L}_m} = {}^{\mathcal{I}}\mathbf{R}_{\mathcal{R}}^\mathsf{T} \left({}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}}^\mathsf{T} (\mathbf{l}_m - {}^{\mathcal{G}}_{\mathcal{G}}\mathbf{p}_{\mathcal{I}}) - {}^{\mathcal{I}}_{\mathcal{I}}\mathbf{p}_{\mathcal{R}} \right), \tag{13}$$

$$d_{\mathbf{l}_m} = \left\| \mathbf{l}'_m \right\| \tag{14}$$

D. Estimator Summary

Summarizing, in our RIO method we propagate the state and its covariance according to Eq. 2 and Eq. 7. The update step of our tightly-coupled EKF consists of three components - the first one makes use of distances to points in the history elements of trails compared to current radar measurements (Eq. 8 and 9), the second one compares distances of persistent landmarks to current radar measurements (Eq. 13 and 14), Finally, the third component employs Doppler velocities as in [15] reduced to the inlier set using 3-point RANSAC as detailed in [6]. For all components, we apply outlier rejection using the chi-squared test.

IV. Experiments

In the following, we outline the setup we used and the experiments we performed to validate our method on a real platform with the data from real flights as well as the results of the evaluation.

A. Experimental Setup

The sensor used for the experiments is a lightweight and inexpensive FMCW radar manufactured by Texas Instruments integrated on an evaluation board AWR1843BOOST, shown attached to the UAV in Fig. 1, equipped with a USB interface and powered with $2.5\,\mathrm{V}$. The frequency spectrum of chirps generated by the radar is between $f_l = 77\,\mathrm{GHz}$ and $f_u = 81\,\mathrm{GHz}$. The Field of View (FoV) is $120\,^\circ$ in azimuth and $30\,^\circ$ in elevation. Measurements are obtained at the rate of $f_m = 15\,\mathrm{Hz}$. The radar is affixed to one extremity of the experimental platform facing forward by a tilt of about $45\,^\circ$ with respect to the horizontal plane as shown in Fig. 1. This improves the velocity readings compared to nadir view while keeping point measurements on the ground and thus at a reasonable distance. For inertial measurements, we use the IMU of the Pixhawk 4 flight controller unit (FCU) with a sampling rate of $f_{si} = 200\,\mathrm{Hz}$. We manually calibrate the transformation between the radar and IMU sensors, which is used as a constant spatial offset in the EKF. The initial navigation states of the filter are set to the ground truth values. N was set to 7. We placed some arbitrary reflective clutter in the scene since the test environment was otherwise a clutter-less clean lab space. No position information from the added objects of any sort was measured or used in our approach other than what the onboard radar sensor perceived by itself.

We use a motion capture system to record the ground truth trajectories. During acquisition, we recorded sensor readings from the IMU and radar together with the poses of the UAV streamed by the motion capture system as ground truth. Our EKF-based RIO is executed offline on the recorded sensor data on an Intel Core i7-10850H vPRO laptop with 16 GB RAM in a custom C++ framework compiled with gcc 9.4.0 at -O3 optimization level. Execution timings for the aforementioned machine are shown in Tab. II and confirm the real-time capability of the implementation.

B. Evaluation

For evaluation of the presented RIO approach, we use the data recorded in an indoor space shown in Fig. 1 during seven manually-controlled UAV flights. The flown trajectories were not pre-planned and included pronounced motions in all three dimensions. One of the executed trajectories can be seen in the Fig. 5. We choose to measure the quality of our estimator using the norm of MAE of position and the final pose drift in percent (without yaw alignment) in order to easily compare against the state-of-the-art. In Fig. 6, one can observe the mean of the ||MAE|| across the flown trajectories. The ||MAE|| increases steadily with the flown distance as the pose drift builds up as expected. For comparison, we show that the ||MAE|| is reduced by a factor of 2 with respect to our previous results presented in [15], and by a factor of 4 with respect to the state-of-the-art shown in [6] where only an FMCW radar and IMU is used and no assumptions on environment are made. The sample based 1σ bounds grow to a value of about $\sigma = 0.25\,\mathrm{m}$. Regarding the final drift, as shown in Tab. I, on average we achieve $0.81\,\%$ on trajectories ranging from $127.5\,\mathrm{m}$ to $175.0\,\mathrm{m}$ against $5.0\,\%$ reported for a $60\,\mathrm{m}$ long flown trajectory in [6] for a similar setup as ours consisting of only FMCW radar and IMU. Authors in [6], [7] also report the final drift values for the same setups, yet additionally augmented with a barometer sensor to reduce the vertical drift - even in this case our results with no additional sensor remain comparable - $0.81\,\%$ (ours) against $0.60\,\%$ [6] and $0.36\,\%$ [7]. Interestingly, the minimum value of the final drift in our experiments - $0.17\,\%$ - is even lower than the lowest value reported in [6], [7] obtained flying the same distance but with additional sensors and much lower excitation in their case. Compared to $3.32\,\%$ final drift for a hand-held trajectory of $116.4\,\mathrm{m}$ with low excitation in [15], our method outperforms it by a factor of 4 in a real UAV flight with high excitation.

TABLE I

METRICS GATHERED ACROSS ALL FLOWN TRAJECTORIES

Trajectory	Length [m]	Norm of MAE at 127 m [m]	Final drift [%]
1	127.5	0.90	**0.17**
2	150.5	1.10	1.62
3	158.4	0.68	0.80
4	160.4	0.71	0.61
5	166.7	0.59	0.78
6	168.1	0.84	0.91
7	175.0	**0.45**	0.81
Average		**0.75**	**0.81**

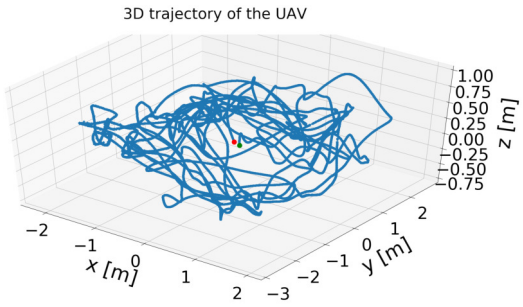

Fig. 5. One of the executed trajectories. As can be seen, we tried to perform pronounced motions in all three dimensions. Green and red dots represent the take-off and landing positions on the white table seen in the Fig. 1.

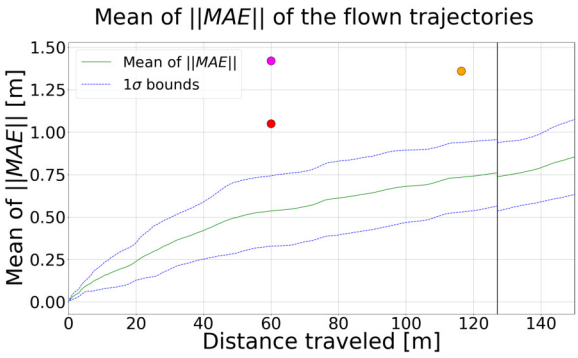

Fig. 6. Mean of the norm of MAE for all flown trajectories. Up to $127\,\mathrm{m}$ the mean of all trajectories is taken. From $127\,\mathrm{m}$ to $150\,\mathrm{m}$ mean of six trajectories is taken (since trajectory 1 ended). Blue dashed lines are the sample based 1σ bounds. For comparison, red and orange dots represent results presented in [15] and the magenta dot represent the state-of-the-art result in [6] for the flown trajectory.

V. CONCLUSIONS

In this paper we presented a tightly-coupled and real-time capable EKF RIO method which builds upon the approach presented in [15] but effectively leverages multi-state and persistent landmarks aspects from the vision community enhanced for the noisy, inaccurate, and sparse radar signals. In the presented framework, for correcting the drift of the IMU, during the update step we exploit lightweight and inexpensive FMCW radar distance measurements to 3D points taken at several time instants in the past, distance measurements to persistent landmarks as well as Doppler velocity measurements. We showed in real-world flight experiments that our method exhibits solid improvements over state-of-the-art in terms of accuracy and that it has execution times suitable for real time control. Moreover, the presented odometry approach makes no assumptions on the environment and can be deployed in GNSS-denied settings. Additionally, given the employed sensor suite it is largely unaffected by conditions deemed challenging for other sensors used in UAV navigation, thus making it promising for applications such as search and rescue operations.

TABLE II

EXECUTION TIMINGS FOR A SINGLE TRAJECTORY

	Average time [ms]			
	Propagation	Update	Duration [s]	Realtime factor
Trajectory 7	0.08	2.07	436	25.36

REFERENCES

[1] P. Fritsche, B. Zeise, P. Hemme, and B. Wagner, "Fusion of radar, lidar and thermal information for hazard detection in low visibility environments," in *2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*. IEEE, 2017, pp. 96–101.

[2] J. Hasch, E. Topak, R. Schnabel, T. Zwick, R. Weigel, and C. Waldschmidt, "Millimeter-wave technology for automotive radar sensors in the 77 ghz frequency band," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 3, pp. 845–860, 2012.

[3] H. Rohling and M.-M. Meinecke, "Waveform design principles for automotive radar systems," in *2001 CIE International Conference on Radar Proceedings (Cat No. 01TH8559)*. IEEE, 2001, pp. 1–4.

[4] M. Schneider, "Automotive radar-status and trends," in *German microwave conference*, 2005, pp. 144–147.

[5] A. Kramer, C. Stahoviak, A. Santamaria-Navarro, A.-A. Aghamohammadi, and C. Heckman, "Radar-inertial ego-velocity estimation for visually degraded environments," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 5739–5746.

[6] C. Doer and G. F. Trommer, "An ekf based approach to radar inertial odometry," in *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2020, pp. 152–159.

[7] ——, "Radar inertial odometry with online calibration," in *2020 European Navigation Conference (ENC)*. IEEE, 2020, pp. 1–10.

[8] ——, "Yaw aided radar inertial odometry using manhattan world assumptions," in *2021 28th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS)*, 2021, pp. 1–9.

[9] D. Barnes and I. Posner, "Under the radar: Learning to predict robust keypoints for odometry estimation and metric localisation in radar," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 9484–9490.

[10] K. Burnett, D. J. Yoon, A. P. Schoellig, and T. D. Barfoot, "Radar odometry combining probabilistic estimation and unsupervised feature learning," *CoRR*, vol. abs/2105.14152, 2021. [Online]. Available: https://arxiv.org/abs/2105.14152

[11] Y. S. Park, Y.-S. Shin, and A. Kim, "Pharao: Direct radar odometry using phase correlation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2617–2623.

[12] S. H. Cen and P. Newman, "Precise ego-motion estimation with millimeter-wave radar under diverse and challenging conditions," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 6045–6052.

[13] R. Aldera, M. Gadd, D. De Martini, and P. Newman, "What goes around: Leveraging a constant-curvature motion constraint in radar odometry," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7865–7872, 2022.

[14] Y. Almalioglu, M. Turan, C. X. Lu, N. Trigoni, and A. Markham, "Milli-rio: Ego-motion estimation with low-cost millimetre-wave radar," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3314–3323, 2020.

[15] J. Michalczyk, R. Jung, and S. Weiss, "Tightly-coupled ekf-based radar-inertial odometry," in *2022 International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Accepted June 2022.

[16] S. I. Roumeliotis and J. W. Burdick, "Stochastic cloning: A generalized framework for processing relative state measurements," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*, vol. 2. IEEE, 2002, pp. 1788–1795.

[17] P. 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, 1992.

[18] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, 2007, pp. 3565–3572.

[19] H. Sommer, I. Gilitschenski, M. Bloesch, S. Weiss, R. Siegwart, and J. Nieto, "Why and how to avoid the flipped quaternion multiplication," *Aerospace*, vol. 5, no. 3, p. 72, 2018.

[20] P. Maybeck, *Stochastic Models, Estimation and Control*, ser. Mathematics in science and engineering. Academic Press, 1979, no. pt. 1. [Online]. Available: https://books.google.at/books?id=eAdRAAAAMAAJ

[21] S. Weiss and R. Siegwart, "Real-time metric state estimation for modular vision-inertial systems," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 4531–4537.

[22] J. Sola, A. Monin, M. Devy, and T. Lemaire, "Undelayed initialization in bearing only slam," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005, pp. 2499–2504.