Tightly-Coupled EKF-Based Radar-Inertial Odometry

Jan Michalczyk[1], Roland Jung and Stephan Weiss

***Abstract*— Multicopter Unmanned Aerial Vehicles (UAV) are small and agile robots with the potential to become prominent in performing autonomous tasks in various Global Navigation Satellite System (GNSS)-denied environments. These environments can potentially be rendered even more challenging due to external factors impairing the robot's perception, such as low or too bright light, permeation with aerosols or smoke. A precondition of autonomous operation, though, is the ability of a robot to accurately localize itself in the surrounding environment. Millimeter-wave Frequency Modulated Continuous Wave (FMCW) radar sensors are resilient to the aforementioned factors while being lightweight, inexpensive and highly accurate. In this paper, we present a Radar-Inertial Odometry (RIO) method for estimating the full 6DoF pose and 3D velocity of a UAV. In an Extended Kalman Filter (EKF) framework, we fuse range measurements and velocity measurements of 3D points detected by an FMCW radar sensor together with Inertial Measurement Unit (IMU) readings. In real experiments we show that our approach enables accurate state estimation of a UAV and that it exhibits improvements over similar existing state-of-the-art method.**

I. Introduction

Increasing UAV autonomy in challenging environments recently attracts substantial interest in the research community. That is because these lightweight and agile platforms have the ability to be successfully used in exploration of environments which are prohibitive to other kinds of robots. Notable examples are environments where GNSS signal is not available or is unreliable and where external conditions prevent cameras from capturing images applicable in the navigation algorithms, such as Visual-Inertial Odometry (VIO). There exist several well-established approaches to address the pose estimation problem of a UAV which fuse measurements from a camera sensor and an IMU. These approaches, however, are limited by known issues of camera sensors like withstanding the lack of sufficient lighting or the excess of it, scarcity of visual features and image blur [1]. These limitations render VIO methods vulnerable in many scenarios, such as environments permeated by smoke or aerosols or while executing very dynamic motions, causing grabbed images to be blurred.

There exist event cameras, whose purpose is to partly mitigate some shortcomings of the traditional cameras, nevertheless, event cameras remain expensive [2], difficult to tune, and are prone to high noise in low-light conditions. LIDAR sensor is also explored in some approaches [3], nevertheless since this modality uses near-visible wavelengths, it is prone to similar drawbacks as cameras in conditions like fog, rain etc.

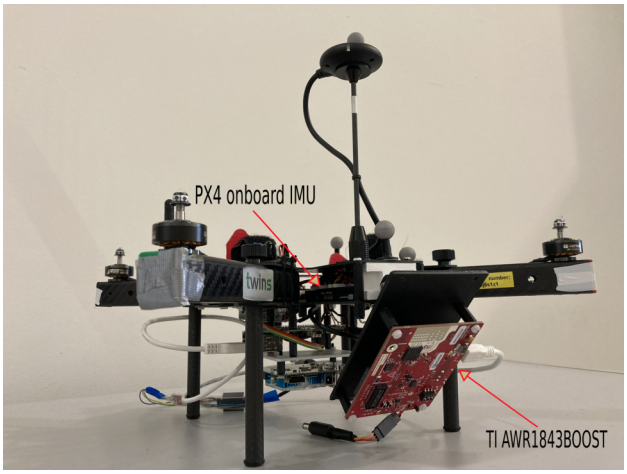

Fig. 1. Experimental platform used in this work with the FMCW radar sensor mounted in its custom-made housing tilted at 45° angle.

Having in mind all the above-mentioned issues, in order to obtain a state estimation system for a UAV it seems reasonable to combine an IMU and a radar sensor. In recent years thanks to using millimeter-wave technology, FMCW radar sensors became considerably miniaturized, lightweight and very accurate. They have been used with success in the automotive applications [4], [5], [6] and the aforementioned technological advances also make them applicable on UAVs.

An FMCW radar sensor is a good compliment of an IMU onboard a UAV. It can mitigate the drawbacks of a camera sensor related to the vulnerability to environmental conditions [7] while at the same time being lightweight and providing comparable or higher measurement rates. Radars provide distance, Doppler velocity, and to certain extent angular measurements of salient points in the environment, which allows reducing the significant pose drift built up by IMU integration. That being said, Radars suffer from problems like e.g., noisy data, ghost objects and multi-path reflections [8], making them non-trivial sensors to be fused with other modalities.

In this paper, we present a novel RIO method which employs stochastic cloning [9] to enable matching of the measured 3D points from the previous radar scan to the ones in the current scan. In addition to these relative distance measurements of matched 3D points, we also use Doppler velocity information measured from all features in the current scan. We fuse all measurements in a tightly-coupled formulation in our EKF setup. The tight coupling

[1]Authors are with the Control of Networked Systems Group, University of Klagenfurt, Austria {jan.michalczyk, roland.jung, stephan.weiss}@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 June/2022, DOI follows ASAP ©IEEE.

enables the incorporation of single distance and velocity measurements in the update step. This property relieves us from any constraints 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 [10] works poorly on noisy and sparse FMCW radar point clouds. 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.

This paper is organized as follows. Section II reviews the related work in the area of RIO. Section III introduces the preliminaries of our work. In subsection III-A we introduce the notation used, the system overview in subsection III-B and brief FMCW radar theory in subsection III-C. In section IV we describe our RIO method. Subsection IV-A outlines how the state vector and covariance matrix of our system are augmented using stochastic cloning in order to enable robust 3D point matching. Subsection IV-B explains our point matching algorithm. Subsequently, subsection IV-C details our distance measurement model. In the subsection IV-D our velocity measurement model is elucidated. We summarize our estimator in subsection IV-E. Experiments conducted in order to validate the proposed method are outlined in Section V. In subsection V-A we report the experimental setup used during the experiments and the subsection V-B presents the results of the evaluation. Finally, we present conclusions in Section VI.

II. Related Work

Robot state estimation methods relying on FMCW radar sensors are becoming more and more prevalent in research. Authors in [11] estimate the ego-velocity of a UAV by fusing readings from IMU and FMCW radar in a nonlinear optimization based framework over a sliding window of past measurements. The presented method is shown to be effective in poorly lit environments as compared with VIO, the emphasis is placed on estimating velocity and not the full 6DoF pose though. In [12], the authors present an EKF based RIO system for a UAV using only instantaneous velocity measurements of detected 3D points, that is, no scan matching is performed. The authors of this work resort to using an additional barometer sensor in order to overcome high vertical drift. In [13], the same authors propose a similar approach (including a barometer for drift reduction), this time with online calibration and using a radar with very wide Field of View (FoV) angle in both azimuth and elevation (130°) which improves the estimation. Methods presented in [14] and [15] use high-resolution, expensive and of considerable size mechanically rotating radar. In the former, a Convolutional Neural Network (CNN) is introduced capable of predicting robust features in radar scans, which are then used to estimate the optimal relative transformation. In the latter method, a direct approach is shown using the entire radar scan as opposed to only extracted features. Namely, the application of Fourier-Mellin transform to Cartesian and log-polar scans is explored to obtain respectively translation and rotation of the robot. The authors in [16] use a rotating radar to estimate the ego-motion, making use of landmarks extraction based on estimating the signal noise statistics and subsequent matching of these landmarks exploiting local geometrical relationships between them within a scan. Size, cost, and power consumption of the rotating radar sensor used in [14], [15], [16] are fairly prohibitive factors limiting its use in UAVs as mentioned in [17]. In [17] an approach to RIO is presented and demonstrated on a slow wheeled mobile robot platform which uses a lightweight and low-cost FMCW radar with no modifications to the environment. This approach employs the Normal Distributions Transform (NDT) for finding the optimal transformation between associated points across subsequent scans. The retrieved pose is fused loosely coupled with an IMU in an Unscented Kalman Filter (UKF). The presented system accomplishes high precision in the order of centimeters. It is, however, evaluated on a flat 2D scene which reduces significantly the dynamics of motion and removes the vertical drift problem as opposed to an application on a UAV.

In this paper, we present a method for real-time capable and accurate RIO, suitable for a UAV using an IMU sensor and a single, small, lightweight, and low-cost FMCW radar providing sparse and noisy 3D point clouds along with Doppler velocities of the detected points. The main contributions of this paper are:

- Tightly-coupled formulation to include *both* radar distance and velocity measurements for IMU integration correction in a C++ EKF framework, allowing accurate 3D velocity and 6DoF pose estimation with nearly a third of the drift compared to state of the art.
- Application of stochastic cloning of past robot poses for formulating an update equation on the accurately measured point *distances* rather than on the full 3D point positions polluted by the highly imprecise azimuth and elevation angular radar measurements.
- Improved 3D point matching across sparse, noisy radar scans in full 6DoF motion allowing ad-hoc point correspondence generation for point-distance based updates with a single past radar pose in the state in contrast to maintaining many 3D point vectors.
- Benchmark comparison against a state-of-the-art method [12] using similar setup in a real-world experiment.

III. Preliminaries

A. Notation

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. 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 ${}^{\mathcal{A}}\mathbf{T}_{\mathcal{B}} = \begin{bmatrix} {}^{\mathcal{A}}\mathbf{R}_{\mathcal{B}} & {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{p}_{\mathcal{B}} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \in \mathrm{SE}(3)$, 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 ${}^{\mathcal{A}}\mathbf{R}_{\mathcal{B}} = \mathbf{R}\left\{{}^{\mathcal{A}}\bar{\mathbf{q}}_{\mathcal{B}}\right\} \in \mathrm{SO}^3$ and ${}^{\mathcal{A}}\bar{\mathbf{q}}_{\mathcal{B}} = \bar{\mathbf{q}}\left\{{}^{\mathcal{A}}\mathbf{R}_{\mathcal{B}}\right\}$ [18]. $\mathbf{I}$ 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.

B. System Overview

In our RIO method we use error-state EKF formulation [19] in which IMU is a core sensor used for the system state propagation. Updates are performed with the FMCW radar measurements, which provide both position and relative radial velocity of reflecting objects. Every time a radar measurement is taken, we augment the state of our EKF filter with the pose of the robot at which the measurement took place using stochastic cloning. Once a subsequent radar measurement is taken, we use the stored pose together with the current one in order to spatially align the radar scans and match the corresponding points across them. Distances to matches are used to form the residual vector in the EKF. 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. 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. 2.

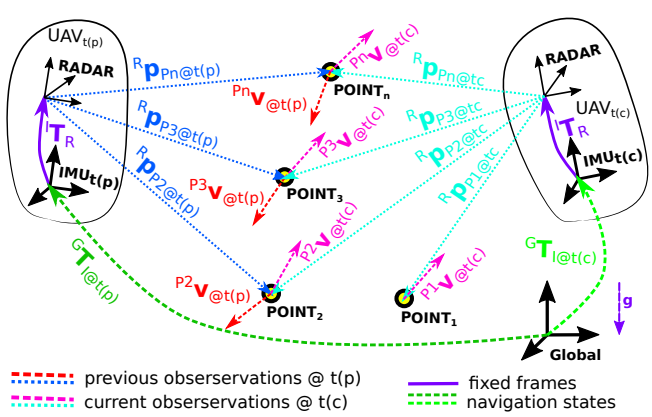

Fig. 2. Two subsequent UAV poses are used in the distance and velocity measurement models. In the velocity measurement model, only the readings from the current pose are used.

C. FMCW Radar Theory

In our setting, we use a millimeter wave FMCW radar emitting a linear chirp signal. In such a case, the range to a reflecting object can be determined from the mixed transmitter and receiver signals. Such mixed signal has the frequency proportional to the time delay of the reflected transmitter signal. Consequently, by applying Fast-Fourier Transform (FFT) over the mixed signal one can infer the range to the reflecting object as:

$$d = \frac{f_d c_0}{2K} \tag{1}$$

Where f_d is the frequency of the mixed signal, c_0 is the speed of light and K is the frequency slope of the linear chirp. By performing 2D FFT over a series of chirps, the relative radial velocity of the object can be retrieved. When multiple receiver antennas are available, calculation of angular information of reflecting objects is possible. Due to the spatial separation of the receiver antennas, the measured signals at each receiving antenna will be phase-shifted,

$$\varphi_{\text{ant}} = \frac{2\pi d_{\text{ant}}}{\lambda} \tag{2}$$

where d_{ant} is the separation distance between receiving antennas. Based on the antenna's geometrical arrangement, either azimuth or elevation angle of objects with respect to the radar can be determined. Using antenna arrays, both angles can be estimated. The angular resolution of a FMCW radar is mostly limited by the number of receiver antennas N_{RX} on the chip. Although FMCW radars are thus capable of providing a 3D position of a reflecting object along with its relative Doppler velocity, the distance to the object and its relative Doppler velocity are the most precise parts of the information in small, light, and cost-efficient FMCW radars. Moreover, the radar that is used in our work provides the intensity of the reflection for each point, which we exploit in our data association algorithm described in subsection IV-B.

IV. Radar-Inertial State Estimation

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

$$\begin{aligned} \mathbf{x}^{\mathsf{T}} &= [\mathbf{x}_{\mathcal{N}}, \mathbf{x}_{\mathcal{C}}] \\ &= \left[\left[{}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}}, {}^{\mathcal{G}}\bar{\mathbf{q}}_{\mathcal{I}}, {}^{\mathcal{G}}\mathbf{v}_{\mathcal{I}}, \mathbf{b}_{\mathbf{a}}, \mathbf{b}_{\boldsymbol{\omega}}\right], \left[{}^{\mathcal{G}}\mathbf{p}_{\mathcal{R}}, {}^{\mathcal{G}}\bar{\mathbf{q}}_{\mathcal{R}}\right]\right] \end{aligned} \tag{3}$$

with the navigation state $\mathbf{x}_{\mathcal{N}}$ and the stochastic clone state $\mathbf{x}_{\mathcal{C}}$ of the IMU pose corresponding to the previous radar measurement as described in Section IV-A. The previous radar measurement is 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 $\mathbf{b}_{\mathbf{a}}$ are the measurement biases of the gyroscope and accelerometer, respectively. ${}^{\mathcal{G}}\mathbf{p}_{\mathcal{R}}$ and ${}^{\mathcal{G}}\bar{\mathbf{q}}_{\mathcal{R}}$ define the pose of the IMU frame corresponding to the last radar measurement $\{\mathcal{R}\}$ with respect to the navigation frame $\{\mathcal{G}\}$. We will use this frame in Sec. IV-B 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 (Sec. IV-C).

The evolution of the state is expressed by the following

differential equations:

$$
\begin{aligned}
{}^{\mathcal{G}}\dot{\mathbf{p}}_{\mathcal{I}} &= {}^{\mathcal{G}}\mathbf{v}_{\mathcal{I}}, \\
{}^{\mathcal{G}}\dot{\mathbf{v}}_{\mathcal{I}} &= {}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}}\left({}_{\mathcal{I}}\mathbf{a}^{\#} - \mathbf{b_a} - \mathbf{n_a}\right) + {}_{\mathcal{G}}\mathbf{g}, \\
{}^{\mathcal{G}}\dot{\mathbf{R}}_{\mathcal{I}} &= {}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}}\left[{}_{\mathcal{I}}\boldsymbol{\omega}^{\#} - \mathbf{b}_{\boldsymbol{\omega}} - \mathbf{n}_{\boldsymbol{\omega}}\right]_{\times}, \\
\dot{\mathbf{b}}_{\mathbf{a}} &= \mathbf{n}_{\mathbf{b_a}}, \dot{\mathbf{b}}_{\boldsymbol{\omega}} = \mathbf{n}_{\mathbf{b}_{\boldsymbol{\omega}}}, {}^{\mathcal{G}}\dot{\mathbf{p}}_{\mathcal{R}} = \mathbf{0}, {}^{\mathcal{G}}\dot{\mathbf{R}}_{\mathcal{R}} = \mathbf{0}
\end{aligned}
\tag{4}
$$

where ${}_{\mathcal{I}}\mathbf{a}^{\#}$ and ${}_{\mathcal{I}}\boldsymbol{\omega}^{\#}$ are the accelerometer and gyroscope measurements of the IMU with a white measurement noise $\mathbf{n_a}$ and $\mathbf{n}_{\boldsymbol{\omega}}$. $\mathbf{n}_{\mathbf{b_a}}$ and $\mathbf{n}_{\mathbf{b}_{\boldsymbol{\omega}}}$ 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 following error state vector from the states defined in Equation (3):

$$
\begin{aligned}
\tilde{\mathbf{x}}^{\mathsf{T}} &= [\tilde{\mathbf{x}}_{\mathcal{N}}, \tilde{\mathbf{x}}_{\mathcal{C}}] \\
&= \left[\left[{}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{I}}, {}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{I}}, {}^{\mathcal{G}}\tilde{\mathbf{v}}_{\mathcal{I}}, \tilde{\mathbf{b}}_{\mathbf{a}}, \tilde{\mathbf{b}}_{\boldsymbol{\omega}}\right], \left[{}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{R}}, {}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{R}}\right]\right].
\end{aligned}
\tag{5}
$$

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 $\tilde{\bar{\mathbf{q}}} = \hat{\bar{\mathbf{q}}}^{-1} \otimes \bar{\mathbf{q}} = \left[1; \frac{1}{2}\tilde{\boldsymbol{\theta}}\right]$, with $\otimes$ and $\tilde{\boldsymbol{\theta}}$ being quaternion product and small angle approximation, respectively.

A. Stochastic Cloning

In order to process relative measurements relating to estimates at different time instances, Roumeliotis and Burdick introduce the concept of Stochastic Cloning (SC) in [9]. 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 (5), $\tilde{\mathbf{x}}_{\mathcal{C}}$ is defined as the error state of the stochastic clone of the IMU pose state $\mathbf{x}_{\mathcal{I}}^{\mathsf{T}} = \left[{}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}}, {}^{\mathcal{G}}\mathbf{q}_{\mathcal{I}}\right]$ and $\mathbf{x}_{\mathcal{O}}^{\mathsf{T}} = \left[{}^{\mathcal{G}}\mathbf{v}_{\mathcal{I}}, \mathbf{b_a}, \mathbf{b}_{\boldsymbol{\omega}}\right]$ are the other states of the navigation state. As cloned state is fully correlated with the IMU pose, it leads to the following stacked/augmented covariance matrix of the corresponding error-state

$$
\tilde{\mathbf{x}} = \begin{bmatrix} \tilde{\mathbf{x}}_{\mathcal{I}} \\ \tilde{\mathbf{x}}_{\mathcal{O}} \\ \tilde{\mathbf{x}}_{\mathcal{C}} \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathcal{I}} & \boldsymbol{\Sigma}_{\mathcal{I}\mathcal{O}} & \boldsymbol{\Sigma}_{\mathcal{I}} \\ \boldsymbol{\Sigma}_{\mathcal{O}\mathcal{I}} & \boldsymbol{\Sigma}_{\mathcal{O}} & \boldsymbol{\Sigma}_{\mathcal{O}\mathcal{I}} \\ \boldsymbol{\Sigma}_{\mathcal{I}} & \boldsymbol{\Sigma}_{\mathcal{I}\mathcal{O}} & \boldsymbol{\Sigma}_{\mathcal{C}} \end{bmatrix}
\tag{6}
$$

with $\boldsymbol{\Sigma}_{\mathcal{N}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathcal{I}} & \boldsymbol{\Sigma}_{\mathcal{I}\mathcal{O}} \\ \boldsymbol{\Sigma}_{\mathcal{O}\mathcal{I}} & \boldsymbol{\Sigma}_{\mathcal{O}} \end{bmatrix}$ being the 15×15 uncertainty of the navigation state $\tilde{\mathbf{x}}_{\mathcal{N}}$, and $\boldsymbol{\Sigma}_{\mathcal{C}} = \boldsymbol{\Sigma}_{\mathcal{I}}$ being the 6×6 uncertainty of the cloned IMU pose error state $\tilde{\mathbf{x}}_{\mathcal{I}}$.

The cloned pose does not evolve with time, meaning no state transition (i.e., $\boldsymbol{\Phi}_{\mathcal{C}}^{k+1|k} = \mathbf{I}$) and no process noise (i.e., $\mathbf{G}_{\mathcal{C}}^{k+1|k} = \mathbf{0}$) is applied, while the original state estimate propagates as usual. From this, the error state propagation can be derived as

$$
\begin{aligned}
\tilde{\mathbf{x}}^{k+1} &= \boldsymbol{\Phi}^{k+1|k}\tilde{\mathbf{x}}^{k} + \mathbf{G}^{k+1|k}\mathbf{w}^{k}, \\
\begin{bmatrix} \tilde{\mathbf{x}}_{\mathcal{N}}^{k+1} \\ \tilde{\mathbf{x}}_{\mathcal{C}}^{k+1} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\Phi}_{\mathcal{N}}^{k+1|k} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_{\mathcal{C}}^{k+1|k} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_{\mathcal{N}}^{k} \\ \tilde{\mathbf{x}}_{\mathcal{C}}^{k} \end{bmatrix} + \begin{bmatrix} \mathbf{G}_{\mathcal{N}}^{k+1|k} \\ \mathbf{G}_{\mathcal{C}}^{k+1|k} \end{bmatrix} \mathbf{w}^{k} \\
&= \begin{bmatrix} \boldsymbol{\Phi}_{\mathcal{N}}^{k+1|k} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_{\mathcal{N}}^{k} \\ \tilde{\mathbf{x}}_{\mathcal{C}}^{k} \end{bmatrix} + \begin{bmatrix} \mathbf{G}_{\mathcal{N}}^{k+1|k} \\ \mathbf{0} \end{bmatrix} \mathbf{w}^{k}
\end{aligned}
\tag{7}
$$

with the linearized state transition matrix $\boldsymbol{\Phi}$ and the linearized perturbation matrix $\mathbf{G}$ computed as explained by Weiss in [20] or related work. The full error-state uncertainty of Equation (6) can then be propagated as

$$
\begin{aligned}
\boldsymbol{\Sigma}^{k+1} &= \boldsymbol{\Phi}^{k+1|k}\boldsymbol{\Sigma}^{k}(\boldsymbol{\Phi}^{k+1|k})^{\mathsf{T}} + \mathbf{G}^{k+1|k}\mathbf{Q}^{k}(\mathbf{G}^{k+1|k})^{\mathsf{T}} \\
&= \begin{bmatrix} \boldsymbol{\Sigma}_{\mathcal{N}}^{k+1} & \boldsymbol{\Phi}_{\mathcal{N}}^{k+1|k}\boldsymbol{\Sigma}_{\mathcal{N}\mathcal{C}}^{k}\mathbf{I} \\ \mathbf{I}\boldsymbol{\Sigma}_{\mathcal{C}\mathcal{N}}^{k}(\boldsymbol{\Phi}_{\mathcal{N}}^{k+1|k})^{\mathsf{T}} & \boldsymbol{\Sigma}_{\mathcal{C}}^{k} \end{bmatrix}
\end{aligned}
\tag{8}
$$

with $\mathbf{I}$ being the identity matrix (since cloned states do not evolve in time), $\mathbf{Q}$ being the discretized process noise matrix, $\boldsymbol{\Sigma}_{\mathcal{C}\mathcal{N}}^{k} = (\boldsymbol{\Sigma}_{\mathcal{N}\mathcal{C}}^{k})^{\mathsf{T}}$ the cross-covariance between the navigation error-state and the stochastic clone error-state, and $\boldsymbol{\Phi}_{\mathcal{N}}^{k+1|k}$ the error-state transition matrix of the navigation error-state $\tilde{\mathbf{x}}_{\mathcal{N}}$. This propagation allows us to rigorously reflect the cross-correlations between the cloned state and the evolved state in our error-state formulation. The above described formalism enables us to correctly use the state variables in order to align the previous radar scan to the current one prior to point matching.

B. 3D Point Matching

In order to estimate the distance to detected points using our measurement model, we need to perform point matching between the current and the previous radar scan aligned to the current UAV pose. This is roughly following the idea of [21] in order to avoid tracking 3D points in the state vector. With a single past pose in the state vector, we can find point correspondences in an ad-hoc fashion, as follows.

We base our point matching algorithm on work described in [16] and [17] for 2D ground vehicle setups and extend it to our 3D UAV setting. Having two consecutive radar scans which are aligned using the pose information stored in the state vector, as the first step, we solve the linear sum assignment problem using the Munkres algorithm [22]. We pose the problem as follows:

$$
min \sum_{i} \sum_{j} C_{i,j} X_{i,j}
\tag{9}
$$

Where X is a boolean matrix where $X_{i,j} = 1$ iff row i is assigned to column j. Constraints of the problem are such that each row is assigned to at most one column and each column to at most one row. Entries of the C matrix are computed as Euclidean distances between all points ${}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}}$ from a previous radar scan at time instance t_p and from a current radar scan at t_c:

$$
C_{i,j} = \| {}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_i}^{t_c} - {}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_j}^{t_p} \|
\tag{10}
$$

In the second step, using the proposed potential matches from the previous step, we build a matrix S of scores where each entry is computed as:

$$s_{i,j} = \frac{1}{1 + \| {}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_i}^{t_c} - {}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_j}^{t_p} \|} \tag{11}$$

unless the value of reflection intensity in the current scan is below a certain threshold or the Euclidean distance is above certain maximum threshold. If either of the two aforementioned conditions holds true, the entry is set to $s_{i,j} = 0$. In the third step, a greedy search is performed on the pairs of points whose corresponding entries in S are non-zero. If a point in the previous scan has more than one candidate for a match in the current scan, then, from among the candidates, we choose the one which minimizes the following expression:

$$d_{i,j} = | \sum_k \| {}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_i}^{t_c} - {}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_k}^{t_c} \| - \sum_k \| {}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_j}^{t_p} - {}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_k}^{t_p} \| | \tag{12}$$

Where ${}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_k}^{\{t_c,t_p\}}$ are already matched points in the current and previous scans respectively. This idea has been exploited in [16] and relies on the fact that subsequent radar scans should ideally keep the relative arrangements between the constituting points. The result of this 3D point matching part is a set of 3D point correspondences (see Figure 3) between a previous and the current radar scan. We can then compare the current radar measurements with the estimated distance value computed from the previous radar scan and the estimated state variables as described in the following section.

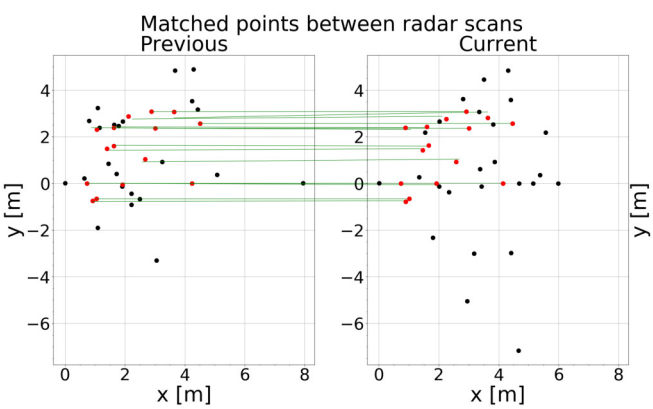

Fig. 3. Matched points between subsequent scans projected onto the x-y plane. Our matching algorithm proves to be robust in the face of outliers and sparse, noisy 3D pointclouds.

C. Distance Measurement Model

In order to estimate the distances to the matched 3D points in the current scan, we transform the corresponding 3D matches ${}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_j}^{t_p}$ from the previous radar scan at time instance t_p to the current radar reference frame, considering the robot's spatial evolution:

$$\begin{aligned} {}^{\mathcal{R}}\mathbf{p}'^{t_p}_{\mathcal{P}_j} = & {}^{\mathcal{I}}\mathbf{R}_{\mathcal{R}}^{\mathsf{T}} \Big(-{}^{\mathcal{I}}_{\mathcal{I}}\mathbf{p}_{\mathcal{R}} + ({}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}}^{t_c})^{\mathsf{T}} \big(-{}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}}^{t_c} + \\ & {}^{\mathcal{G}}\mathbf{p}_{\mathcal{I}}^{t_p} + {}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}}^{t_p} \left({}^{\mathcal{I}}_{\mathcal{I}}\mathbf{p}_{\mathcal{R}} + {}^{\mathcal{I}}\mathbf{R}_{\mathcal{R}} {}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_j}^{t_p} \right) \big) \Big) \end{aligned} \tag{13}$$

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 previous and current radar scans at t_p and t_c, respectively, with respect to the navigation frame $\{\mathcal{G}\}$.

Note that, at this point, we could already formulate a measurement for the matched 3D point in the past with the currently measured one. However, as mentioned in Sec. III-C, low-cost FMCW radars have fairly precise measurements of the object's distance and Doppler velocity, but heavily lack of precision in azimuth and elevation. Thus, we transform the 3D point from Cartesian space to Spherical coordinates and only use the most informative dimension, the distance. Additional measurement formulations for azimuth and elevation could be included with higher measurement uncertainty. The low information versus added complexity and the non-Gaussian noise distribution in these dimensions are, however, arguments to not include them in our RIO framework.

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

$$d_{\mathcal{P}_j} = \left\| {}^{\mathcal{R}}\mathbf{p}'^{t_p}_{\mathcal{P}_j} \right\| \tag{14}$$

where $d_{\mathcal{P}_j}$ is the distance to a single matched 3D point ${}^{\mathcal{R}}\mathbf{p}'^{t_p}_{\mathcal{P}_j}$ in the previous radar scan at t_p aligned to the current radar pose at t_c. Since this measurement relates to states from pastime instances, stochastic cloning is necessary as introduced in Section IV-A.

D. Velocity Measurement Model

In order to estimate the velocities of the detected radar 3D points ${}^{\mathcal{R}}\mathbf{v}_{\mathcal{P}_i}$ in the current scan at t_c, we transform the current robot ego-velocity from the IMU frame into the current radar frame and subsequently project it onto the direction vector pointing towards the corresponding 3D point. This is expressed by the following measurement model:

$$\begin{aligned} {}^{\mathcal{R}}\mathbf{v}_{\mathcal{P}_i} = & \frac{\mathbf{r}^{\mathsf{T}}}{\|\mathbf{r}\|} \big({}^{\mathcal{I}}\mathbf{R}_{\mathcal{R}}^{\mathsf{T}} {}^{\mathcal{G}}\mathbf{R}_{\mathcal{I}}^{\mathsf{T}} {}^{\mathcal{G}}\mathbf{v}_{\mathcal{I}} + \\ & \mathbf{R}_{\mathcal{R}}^{\mathsf{T}} \left({}_{\mathcal{I}}\boldsymbol{\omega} \times {}^{\mathcal{I}}_{\mathcal{I}}\mathbf{p}_{\mathcal{R}} \right) \big) \end{aligned} \tag{15}$$

where $\mathbf{r} = {}^{\mathcal{R}}\mathbf{p}_{\mathcal{P}_i}$ is the 3D point detected in the current scan, ${}_{\mathcal{I}}\boldsymbol{\omega}$ is the current angular velocity of the IMU in the IMU frame, and ${}^{\mathcal{G}}\mathbf{v}_{\mathcal{I}}$ is the current linear velocity of the IMU in the navigation frame. In order to reject outliers, we apply a chi-squared test to each measurement's residual, in which we check if the Mahalanobis distance corresponding to the residual is contained within the interval defined by the thresholds associated with a chosen percentile of the χ^2 distribution.

E. Estimator Summary

In summary, our EKF-based RIO approach consists of Eq. 4 and Eq. 8 to propagate the state and its covariance using the IMU measurements. We then use a tightly-coupled formulation to compare the distances of matched features with current radar distance measurements using Eq. 13 and

Eq. 14, and also include in a tightly-coupled fashion the velocity information the radar sensor provides using Eq. 15 to correct IMU integration errors. The inclusion of both the point distance measurements and current point velocity information in a tightly-coupled fashion is key to the improved performance of our approach compared to state-of-the-art methods.

Although used in the position and velocity updates (Eq. 13 and Eq. 15), we do not keep 3D points in the state vector. This idea is borrowed from [21] where 3D points are triangulated from images on-the-fly without inclusion in the state vector. Our adaptation to RIO and highly simplified implementation of this idea suffers from reduced estimation consistency, but results in less complexity. A thorough analysis of the statistical impact of this simplified implementation can be tackled in future work.

V. Experiments

The above described approach enables a simple, yet computationally efficient RIO method. In the following, we test our method on a real platform with real data.

A. Experimental Setup

The sensor used for the experiments is the lightweight and inexpensive $77\,\mathrm{GHz}$ multichannel millimeter-wave FMCW radar transceiver 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 radio frequency (RF) signals propagate in a FoV of $120\,^\circ$ in azimuth and $30\,^\circ$ in elevation. Measurements are obtained at the rate of $f_m = 20\,\mathrm{Hz}$. The radar is attached 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 with a random offset drawn from the states initial uncertainties as listed in Table I. The above described platform is moved in a hand-held manner across a spacious room, as shown in Fig. 4, repeatedly performing five times the same rectangular-shaped trajectory of approximate dimensions of slightly more than $[4.5\,\mathrm{m} \times 5.5\,\mathrm{m}]$ traversing the total distance of $116.4\,\mathrm{m}$. The scene was augmented with some arbitrary reflective clutter since the test environment was otherwise a clutter-less clean lab space. No position information from the added objects of any sort was 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. Our EKF-based RIO is executed offline but at real-time speed on the recorded sensor data on an Intel Core i7-10850H vPRO laptop with $16\,\mathrm{GB}$ RAM in a custom C++ framework.

TABLE I

Initial standard deviation of the navigation states

	${}^{\mathcal{G}}_{\mathcal{G}}\mathbf{p}_{\mathcal{I}}$	${}^{\mathcal{G}}_{\mathcal{G}}\mathbf{v}_{\mathcal{I}}$	${}^{\mathcal{G}}\mathbf{q}_{\mathcal{I}}$	${}_{\mathcal{I}}\mathbf{b}_a$	${}_{\mathcal{I}}\mathbf{b}_\omega$
$\boldsymbol{\sigma}^0$	$0.32\,\mathrm{cm}$	$0.32\,\mathrm{cm/s}$	$0.1\,\mathrm{rad}$	$0.71\,\mathrm{m/s^2}$	$0.1\,\mathrm{rad/s}$

Fig. 4. Indoor space where experiments were performed with reflecting clutter scattered on the scene. The objects were placed randomly and no global position (nor attitude) information of any sort was used in our approach.

B. Evaluation

We evaluate our RIO approach with the data recorded in an indoor space equipped with a motion capture system. Ground truth trajectories as well as the estimated ones can be seen in Fig. 5 to Fig. 8. On these plots, one can observe drift of the estimate versus the ground truth in position and yaw, since these four dimensions are unobservable in a RIO framework. The amount of drift is a direct measure of quality for a given approach.

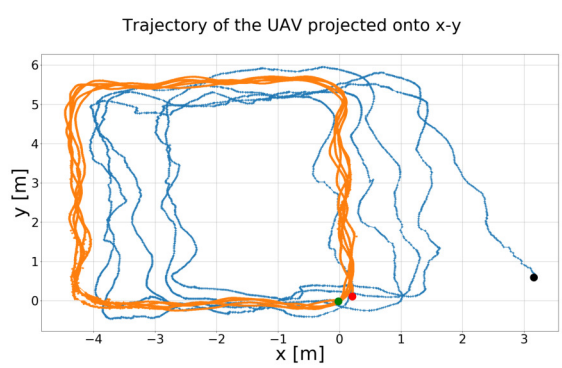

Fig. 5. Top view of the UAV 3D trajectory. The true trajectory is plotted in orange and the estimated one in blue. The biggest (angular) drift can be noted after the last turn. Note the coloured dots marking the end and the beginning of trajectories.

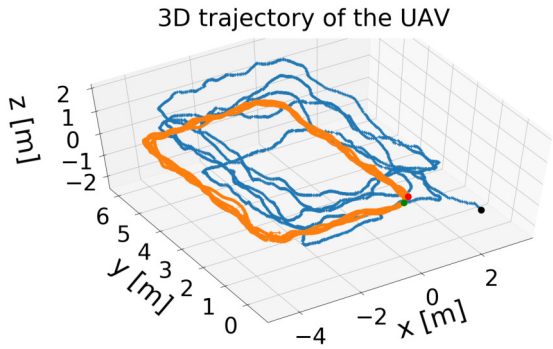

Fig. 6. Side view of the UAV 3D trajectory. The true trajectory is plotted in orange and the estimated one in blue. The total of 116.4 m distance is covered.

Assessment of position, attitude, and velocity tracking is provided on plots 7, 8 and 9 respectively. For the position, one can clearly see the random walk behavior of this unobservable state. On this aspect, note that the provided metrics below are a snapshot of such a random walk (i.e., one of many realizations) – as are the numbers in e.g., [12]. Nevertheless, the vast improvement against state-of-the-art shows that our approach generally has some beneficial aspects.

For the attitude, the drift in yaw is clearly visible. In this dataset we also observe an offset in pitch occurring after the first few seconds and remaining throughout the rest of the run. We assume a slight misalignment of the tracking system reference frame with respect to gravity. This value is observable in a RIO framework and will converge towards a gravity aligned reference frame.

The 3D velocity state is observable and the plot, besides the noticeable jitters also in the ground truth data, does not show a particularly unexpected behavior. The ground truth was computed by numerically differentiating the tracking system's position signal.

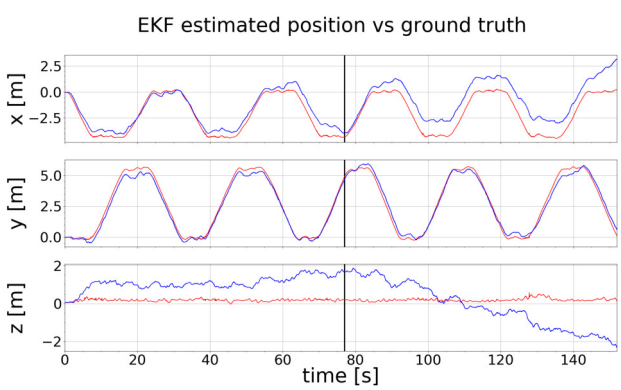

Fig. 7. Estimated position plotted in blue against ground truth plotted in red. Note the thick black line passing through the time instant at which 60 m distance has been traversed (this is for better metrics comparison against state of the art detailed in the text).

On figure 10 we plot the drift from the true trajectory as percent of the traveled distance, drift in meters (norm of the position error) and the Total Traveled Distance (TTD).

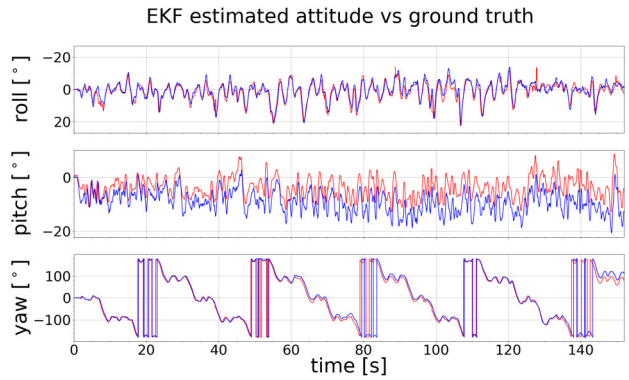

Fig. 8. Estimated attitude plotted in blue against ground truth plotted in red. The offset in pitch angle is assumed to be caused by a slight misalignment of the motion capture system's reference frame and the gravity.

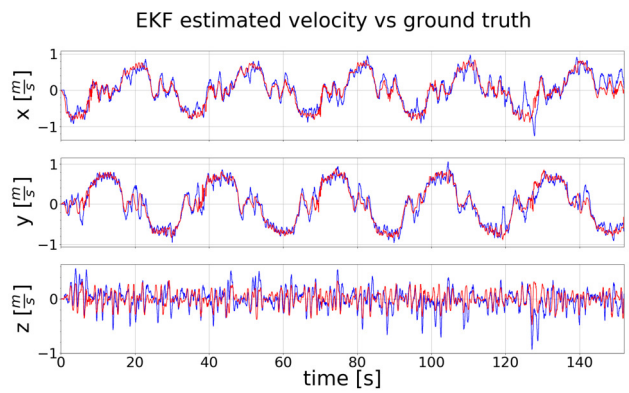

Fig. 9. Estimated velocity plotted in blue against ground truth plotted in red. We compute the ground truth velocity by numerically differentiating the motion capture system's position.

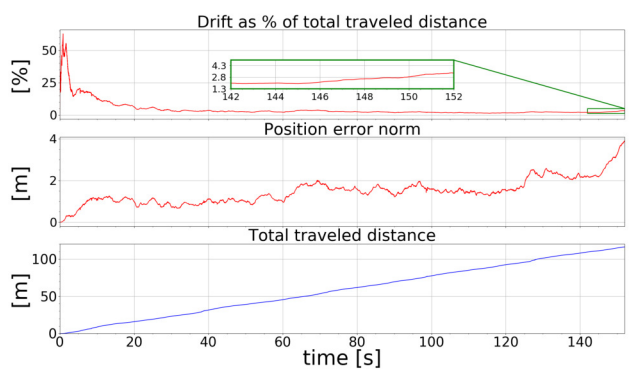

Fig. 10. Drift as % of the TTD, position error norm and the TTD. The drift (position error norm) remains below 4 m throughout the executed trajectory and reaches 3.32 % at the end of the TTD.

The first $t_{init} = 8\,\mathrm{s}$ of the experiment was the initialization period during which agile motions were performed in-place in order to initialize the filter before executing the trajectory. All values are plotted after this initialization phase. We choose to compare our work to the state-of-the-art approach in [12] in which the authors use the same radar sensor to provide metrics for their loosely coupled radar-velocity based RIO approach using indoor hand-held data. In the following, we show the same metrics underlining the benefit of a) tight

coupling and b) use of both distance and velocity information from the radar sensor. Note that in the same work, the authors also included a barometric pressure sensor in their RIO to further reduce drift. This would, however, be a different comparison.

Firstly, we evaluate final drift values. We can note on the Fig. 10 that over the TTD of $116.4\,\mathrm{m}$ we achieve a final drift below $4\,\mathrm{m}$. Authors in [12] in their hand-held experiment achieve drift close to $4\,\mathrm{m}$ (or about $6.67\,\%$) over the trajectory of $60\,\mathrm{m}$, which is roughly half of the distance covered in our experiment. Exact numbers for our approach are $3.86\,\mathrm{m}$ ($3.32\,\%$ for TTD equal to $116.4\,\mathrm{m}$) and $1.54\,\mathrm{m}$ ($2.56\,\%$ for TTD equal to $60\,\mathrm{m}$). Next, we compare the norm of Mean Absolute Error (MAE) of position and velocity for TTD of $60\,\mathrm{m}$. For our approach, these numbers are $1.05\,\mathrm{m}$ and $0.38\,\frac{\mathrm{m}}{\mathrm{s}}$ respectively, against $1.95\,\mathrm{m}$ and $0.14\,\frac{\mathrm{m}}{\mathrm{s}}$ for authors in [12]. At the TTD of $116.4\,\mathrm{m}$ the norm of MAE of position equals to $1.36\,\mathrm{m}$ and is still lower than the one from [12] at $60\,\mathrm{m}$ despite the inherent drift of the unobservable position in RIO approaches. The norm of MAE of velocity for our method at TTD of $116.4\,\mathrm{m}$ equals to $0.36\,\frac{\mathrm{m}}{\mathrm{s}}$. The higher MAE in velocity compared to state of the art is an interesting fact: the velocity states are fully observable and should converge well. For ground truth velocity generation, we used the tracking system's position at $100\,\mathrm{Hz}$ and used a median filtered, delta-time scaled difference of this signal. This has highly limited precision and may have contributed to the relatively large MAE in velocity.

VI. Conclusions

In this paper we presented a tightly coupled EKF based approach to RIO in which we fuse IMU readings with both velocity and distance measurements to 3D points detected by lightweight and inexpensive FMCW radar. We make use of the past IMU pose and a rigid body assumption such that we can generate several point correspondences between two radar scans in an ad-hoc fashion. This requires only maintaining (through stochastic cloning) a 6DoF past pose in the state vector compared to tracking many 3D point vectors. In this context, we showed that our improved matching method can cope with noisy and sparse radar point clouds and generate reliably point correspondences between two scans. With this, we showed that using both distance and velocity measurements, we accomplish accurate 6D pose and 3D velocity estimation of a mobile platform (in this case a hand-held UAV). In particular, the use of the accurately measured distance information to a point is beneficial as it does not include the point's azimuth and elevation angular information which is generally very poorly measured by low-cost FMCW radar. Moreover, we showed that using our method, we reduce the position drift compared to similar state-of-the-art approach. Last but not least, our method is applicable in a variety of environments where potentially GNSS systems are unavailable and vision sensors, commonly used for UAV navigation, cannot be relied upon.

References

[1] A. Pretto, E. Menegatti, M. Bennewitz, W. Burgard, and E. Pagello, "A visual odometry framework robust to motion blur," in *2009 IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 2250–2257.

[2] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam," *The International Journal of Robotics Research*, vol. 36, no. 2, pp. 142–149, 2017.

[3] J. Zhang and S. Singh, "Visual-lidar odometry and mapping: Low-drift, robust, and fast," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 2174–2181.

[4] 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.

[5] 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.

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

[7] 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.

[8] C. X. Lu, S. Rosa, P. Zhao, B. Wang, C. Chen, J. A. Stankovic, N. Trigoni, and A. Markham, "See through smoke: robust indoor mapping with low-cost mmwave radar," in *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, 2020, pp. 14–27.

[9] 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.

[10] 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.

[11] A. Kramer, C. Stahoviak, A. Santamaria-Navarro, A.-A. Agha-Mohammadi, 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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

[20] 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.

[21] 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.

[22] J. Munkres, "Algorithms for the assignment and transportation problems," *Journal of the society for industrial and applied mathematics*, vol. 5, no. 1, pp. 32–38, 1957.