Radar-Inertial State-Estimation for UAV Motion in Highly Agile Manoeuvres

Jan Michalczyk[1], Christian Schöffmann, Alessandro Fornasier, Jan Steinbrener and Stephan Weiss

***Abstract*— Multicopter Unmanned Aerial Vehicles (UAV) are known for their high agility and aggressive manoeuvres. Despite significant advances in state estimation for such vehicles with multiple sensors, their accurate state estimation in highly agile manoeuvres is still a challenge in the research community. In this paper, we present a radar-inertial based method for estimating the full 6D pose and 3D velocity of a UAV including sensor extrinsics and Inertial Measurement Unit (IMU) intrinsics. In an Extended Kalman Filter (EKF) framework, we fuse range measurements of corner reflectors detected by a Frequency Modulated Continuous Wave (FMCW) radar sensor together with IMU readings. Our tightly coupled fusion approach and the high-frequency state correction together with the inherent benefits of radar sensors (e.g. resilience to aerosols, light changes, etc) enables tracking of highly aggressive trajectories in real experiments which are shown to be particularly challenging for a state of the art Visual-Inertial Odometry (VIO) approach we compare against.**

I. Introduction

Increasing UAV autonomy in challenging environments is currently a widely researched topic. These lightweight, rapidly-moving platforms have the potential to serve in exploration of environments which are inaccessible to other kinds of robots. Of particular interest are environments where Global Navigation Satellite System (GNSS) signal is not available and where cameras are not able to capture images useful for the navigation algorithms such as VIO. Also, in order to robustly navigate through the environment with agility, UAVs need fast and reliable state information for the closed-loop motion control. There exist a number of approaches to tackle the pose estimation problem of a UAV which fuse measurements from a camera sensor and an IMU. These approaches, however, suffer from known pitfalls of camera sensors like coping with a lack of lighting, lack of features, image blur [1] and are limited by the rate at which images are acquired by the camera sensor. These limitations render VIO methods vulnerable in many scenarios such as executing aggressive trajectories. There exist event cameras, whose purpose is to mitigate some of the shortcomings of the traditional cameras, nevertheless, event cameras remain expensive [2], difficult to tune, and are prone to high noise in low-light conditions.

[1]All authors are with the Institute of Smart Systems Technologies, University of Klagenfurt, Austria {jan.michalczyk, christian.stetco}@aau.at, {alessandro.fornasier, jan.steinbrener, 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 April/2022, DOI follows ASAP ©IEEE.

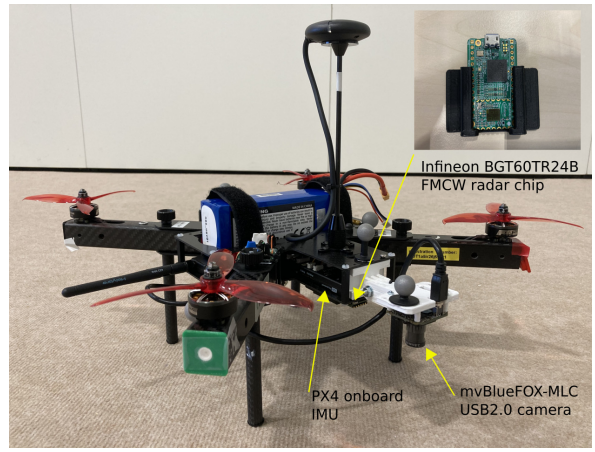

Fig. 1. Experimental platform used in this work and the FMCW radar sensor mounted in its custom-made housing.

Given the previously enumerated issues, a viable approach to estimate the pose of a UAV is to combine an IMU with a radar sensor. Radar sensors have been successfully used in automotive industry [3], [4], [5] for the past few years. Potentially, using a radar onboard a UAV can alleviate drawbacks of a camera sensor, since radar can acquire data at rates several times faster than a camera while at the same time being lightweight and insensitive to the lighting conditions of the environment in which it operates. In addition, radars are able to cope with fog or smoke in the air [6], making them more robust in such scenarios. That being said, they suffer from problems like e.g. noisy data, ghosts objects and multipath reflections [7], making them non-trivial sensors to be fused with other modalities.

In this paper, we present a method which uses a tightly-coupled EKF formulation in which we fuse measurements from an IMU with radar range measurements of targets in the environment. This approach enables fast and accurate provision of the UAV state estimate even in scenarios where VIO suffers from both low rate (relative to the trajectory agility) correction information and image blur adversely affecting the state estimates. Also, a tightly-coupled formulation allows for state updates having sparse measurements, that is, contrarily to loosely-coupled formulations, ours does not require a prior 3D triangulation step for which distances to at least three features must be measured simultaneously.

This paper is organized as follows. Section II reviews related work in the area of radar-based state estimation. Section III gives details on basic FMCW radar theory and our state estimation method. Subsection III-A introduces the system overview. Subsection III-B briefly describes the

concept of FMCW radars. Subsection III-C briefly describes how the radar signal is processed in order to be used in the estimation procedure. Subsection III-D describes how the range measurements from the radar sensor are matched to the reflectors at every iteration. Subsection III-E explains how the tightly-coupled EKF is designed. Experimental validation of the proposed method and results are outlined in Section IV. In subsection IV-A we report the experimental setup we used during the experiments. Subsection IV-B presents the results of the evaluation. Finally, we draw conclusions in Section V.

II. Related Work

Several approaches are found in the recent literature where radar is used to estimate the motion of a mobile robot. In [8] the ego-velocity is estimated from an IMU and a millimeter-wave FMCW radar measurements by solving a nonlinear optimization problem over a sliding window of previous measurements. The estimated ego-velocity is then compared against a VIO-obtained estimate showing the usefulness of the method in environments with limited lighting. This method, however, while showing good results estimating the ego-velocity, does not provide the full 6D pose of the robot nor any extrinsic calibration capability. The rate at which ego-velocity is estimated is not reported, the radar rate was only 10 Hz. In [9] a system based on a Convolutional Neural Network (CNN) is presented, which allows the prediction of robust features in radar images which are then used to compute the pose between two subsequent scans by matching the predicted features and estimating the optimal transform between them. This system uses a high-resolution rotating radar though. As opposed to the above mentioned feature-based method, in [10] a direct method is presented, that is, one where a whole radar scan information is taken into account instead of features. Namely, a Fourier-Mellin transform is applied onto radar cartesian and log-polar images to estimate respectively rotation and translation of the robot. Again though, a high-end rotating radar is used for the state estimation. Both methods operate at frequencies of 28.5 Hz and 10 Hz respectively which might be a limiting factor in performing high-speed manoeuvres with UAVs. In [11] again a rotating FMCW radar is used in a framework with a CNN which calculates embeddings based on radar scans in order to recognize places used subsequently to localize the robot on a known map by computing similarities between the embeddings. Another application of a rotating FMCW radar for ego-motion estimation includes probabilistic landmark extraction and subsequent scan matching [12]. Importantly, the size, cost and power consumption of the rotating radar sensor used in [9], [10], [11], [12] is a heavily limiting factor for its use in UAVs [13]. In [13] an approach to Radar-Inertial Odometry (RIO) is presented on a slow turtlebot which uses a lightweight and low-cost FMCW radar with no modifications to the environment. In this approach the Normal Distributions Transform (NDT) is used for scan matching. Matched scans are used for pose estimation which is fused with IMU in an Unscented Kalman Filter (UKF). The presented method achieves centimeter precision

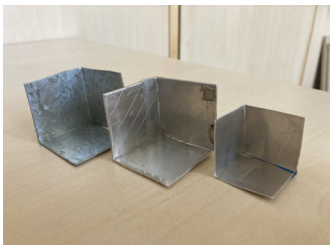

Fig. 2. Corner reflectors used in our experiments.

in indoor environments, it is however not evaluated on fast and aggressive motions for which data associations would be much more challenging and also requires Angle-of-Arrival (AoA) information from the radar which is known to have very poor resolution on low-cost FMCW arrays [7], [14]. In [15] authors show an EKF-based method which fuses ego-velocity estimated from a single millimeter-wave FMCW radar scan using linear least squares with IMU measurements to estimate the state of a UAV. However, to compensate for the large height-measurement error the authors add a barometric altitude sensor to the system. Also, aggressive manoeuvres are not in the scope of the evaluation of that approach. In [16] an approach to VTOL is presented using FMCW radar measurements to known targets fused with the IMU. Fusion is achieved within a loosely-coupled EKF making use of the noisy 3D point measurements instead of the more accurate distance readings. In [17] authors present a radar-based method for 3D localization in landing situations, nevertheless the radar used in this work is a bulky cooperative sensor.

Our work contributes a method for fast and accurate state estimation suitable for a rapidly moving UAV using a single small, lightweight, and low-cost FMCW radar measuring ranges to so-called *corner reflectors* [18] in the environment and an IMU sensor. The approach uses reflectors at known positions and self-calibrates system extrinsics, IMU extrinsics, and estimates the 6DoF pose with 3D velocity. We show that with this method we can outperform VIO when sharp and aggressive motions are executed. Moreover, the presented approach can be used in environmental conditions completely prohibitive for VIO like fog or lack of light. We evaluate our approach experimentally and compare against the state-of-the-art VIO algorithm in [19], using the platform shown in Fig. 1. Although VIO is generally thought of as an odometry approach where no external global features are used for positioning, in our work we use a VIO with *persistent features* that render it locally non-drifting. Persistent features are robust features seen in the environment and which thus are added to the state vector. Consequently, if the motion is such that these features are continuously in the field of view, they (locally in this area) eliminate the drift. The fact that in our experiments the UAV operates in the same restricted area allows the persistent features to be always in the field of view allowing a fair comparison against our RIO approach. Comparison with the existent RIO approaches used on UAVs such as [20], [15], [21] is not as meaningful

as with a VIO that includes persistent features since the drift is always present in those, making the contrast unfair.

III. Methodology

A. System overview

We consider our approach a tightly coupled approach since we use the one-dimensional distance measurement from the onboard radar sensor to a radar target as information for IMU integration correction. We do not pre-process several distance measurements from several targets to triangulate first a 3D position to only then use this result as position correction. Compared to the pre-triangulation method, our tightly coupled approach has several advantages: first, a quick non-linear observability analysis shows that all motion states including IMU biases are locally observable if the system is excited in acceleration and angular velocity while observing only one target (intuition to this fact can be gained from [22]). Thus, the radar does not need to observe all targets at once like in the triangulation pre-processing approach. Note that, at least sequentially, it needs to observe at least two different targets to eliminate any gauge freedom. Second, omitting the pre-processing step from the distance measurement to a 3D position measurement reduces the distortion of the noise characteristics that may otherwise impact the EKF assumptions on Gaussian distributions. Third, we can process measurements as they arrive and do not need any sort of synchronization between the targets. Fourth, the only geometric condition to avoid singularities (in this case to eliminate the gauge freedom in 3D position and yaw) is that at least two targets have different positions in a direction perpendicular to gravity (i.e. in the xy-plane).

Our estimator setup is such that at least two of the targets need to be placed at known location to eliminate said gauge freedom (3D position and yaw), positions of additional targets can seamlessly be integrated in the estimation process. For the initialization of additional targets online, the approach in [23] could be adapted to use initial distance measurements from a newly observed target. That said, in this paper we only focus on the radar-inertial estimation of the motion states as proof of concept.

The targets generate a clear signal in the radar receiver. We perform a Fast Fourier-Transform to extract the distance to each target and then use this information in the EKF framework. The data association is done by matching the measured range closest to the expected range calculated from the current vehicle pose and a given target location. The following subsections detail each step further.

B. FMCW Radar Theory

FMCW radars rely on frequency-modulated continuous waves emitted into free space. In case of linear sweeps, such signals are known as linear chirp signals described by

$$x_T(t) = Ae^{j\left(\omega t + \pi S t^2\right)} + A^* e^{-j\left(\omega t + \pi S t^2\right)} \tag{1}$$

where $(\cdot)^*$ denotes the complex conjugate, A is the signal amplitude, $S = \frac{B}{T_s}$ is known as the chirp rate, $\omega = 2\pi f_T$ is the lower chirp frequency, and B is the bandwidth of the signal. FMCW radar sensors usually transmit multiple chirp signals as described by Eq. (1) sequentially in time. Surrounding objects in the vicinity of the radar sensor reflect the emitted signal and are captured by the receiver antenna. The resulting signal at the receiving element of the radar can be modeled as a time-delayed and damped copy of $x_T(t)$. This time delay is known as the Time-of-Flight (ToF) and is given by $\tau = \frac{2(d+vt)}{c}$ where c is the speed of light, d is the distance to the object, and v is the object's velocity. The received chirp signal is processed in an electronic mixer and low-pass filtered giving a signal modeled as

$$x_R(t) = \tilde{A}e^{j\gamma(t)} + \tilde{A}^* e^{-j\gamma(t)} \tag{2}$$

where $\gamma(t) = 2\tau(\omega + 2\pi S t - \pi S \tau)$. For digital processing, the receiver signal $x_R(t)$ is discretized using an Analog-to-Digital Converter (ADC) operating at a sampling rate f_s. Each chirp is represented by a finite sequence with length denoted as number of samples N_s. Changing N_s or N_c affects sensing range d_{max} as well as range and velocity resolution Δr and Δv, respectively. These quantities are summarized by

$$d_{\text{max}} = \frac{N_s c}{4B} \qquad \Delta r = \frac{c}{2B} \qquad \Delta v = \frac{\lambda f_s}{2 N_c N_s} \tag{3}$$

where λ is the wavelength of the electromagnetic wave [24].

C. Radar Signal Processing

Range and velocity can be determined from the so-called instantaneous frequency obtained as the time-derivative of the exponent in Eq. (2), $f_i(t) = \frac{1}{c_0}\left(\frac{\omega}{\pi}v + 2Kd_0 + 4Kvt\right)$. The first term is known as the Doppler shift, the second term only depends on the range, and the third term is the change of the frequency due to movement of the object. By using signal processing techniques such as Fast-Fourier Transform (FFT) processing, these quantities can be extracted from the instantaneous frequency. By using multiple receiver antennas, extraction of spatial information of surrounding objects is achieved. Due to the spatial separation of the receiver antennas, the measured signals at each receiving antenna will be phase shifted,

$$\varphi_{\text{ant}} = \frac{2\pi d_{\text{ant}}}{\lambda} \tag{4}$$

where d_{ant} is the separation distance between receiving antennas. Based on the antenna 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}. Single objects can be estimated very accurately using low number of antennas. However, resolving multiple closely spaced objects remains a difficult task using few receiver antennas

In the proposed work, the receiver signal is convolved with a Hamming window to minimize spectral leakage. Very low and high frequency contributions which are not in the region of interest are filtered out using a bandpass filter. The bandpass filter was designed with cut-off frequencies

mapped to ranges $r_{c,L} = 0.1$ m and $r_{c,H} = 1.7$ m to suppress frequency peaks caused by noise and multipath reflections. After filtering, the signal is thresholded above an experimentally obtained value of $c_{\text{noise}} = 0.15$ and the resulting signal is searched for clusters. The frequency value corresponding to the maximum value of each cluster is returned and mapped to a range value (Fig. 3).

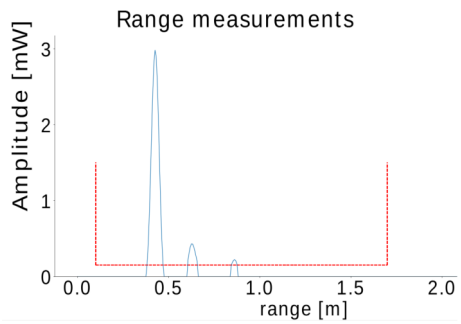

Fig. 3. Range-amplitude plot after performing FFT of the radar raw signal. The dashed red lines depict thresholds assumed for cropping the radar signal in frequency (mapped to range) and for application of the threshold above which targets are identified.

D. Data association

With the output of the processing module we obtained range values of all targets in the field of view of the radar sensor. Compared to obtaining Doppler velocity information or angular information, the range values only need a single FFT pass for calculation. This enables very fast measurements and subsequent propagation correction in the estimator framework. However, this lightweight data processing comes at the difficulty of data association. Each identified peak (i.e. measurement) in the range-amplitude plot (Fig. 3) needs to be associated with a target in the world to derive the necessary correction in the EKF formulation. For this association, we first compute estimates of the ranges between the current vehicle position and the targets from the information we have at the current time step. Then, a greedy search is minimizing the error of a given measurement with the computed possible range estimates. If the previous radar scan already had a similar range measurement, this information is included as a prior. Since the radar scans are very fast compared to the vehicle motion, this element helps speeding up the data association process. The matched range measurement with a target forms then an update information pair for the EKF. If a range measurement differs more than 3 cm to any possible target range, then the measurement is discarded.

E. Tightly coupled Extended Kalman Filter formulation

The EKF framework uses an IMU for the propagation of the state formulation defined by Eq. (5). The states are the position of the IMU/body frame p_{WI} and velocity v_{WI} expressed with respect to the world frame, the orientation of the IMU in the world frame q_{WI}, gyroscopic bias b_ω and accelerometer bias b_a. The 3D translation between the onboard radar senor and the IMU is expressed as the calibration state p_{IS} in the IMU frame. r_W^i represents the i-th corner reflector (i.e. radar target) 3D position in the world frame with $i = 1, 2, ...$ Note that at least two such targets need to be known and fixed to eliminate the gauge freedom. The full state vector X is then defined as follows:

$$X = \left[p_{WI}^T, v_{WI}^T, q_{WI}^T, b_\omega^T, b_a^T, p_{IS}^T, {r_W^i}^T\right]^T \tag{5}$$

The system dynamics of the core states are defined according to [25]. The dynamics and process noise of the calibration state p_{IS} are assumed to be zero because of the rigid body assumption. In our setup, r_W^i are kept fix as known values. The EKF framework uses a regular error state (including error quaternion) definition. Please note that this work uses the Hamilton notation for the Quaternion representation [26].

An estimated range measurement $\hat{z}^i$ from the radar sensor to a target i and the corresponding error using the true measurement z^i from the data association process can then be defined as

$$\hat{z}^i = |r_W^i - (p_{WI} + R(q_{WI})p_{IS})| \tag{6}$$

$$\tilde{z}^i = z^i - \hat{z}^i \tag{7}$$

Where $R(q_{WI})$ rotates a vector from the IMU frame to the world frame using the quaternion q_{WI} and $|\cdot|$ represents the L_2 norm. This renders the update step very efficient for any number of currently visible targets by the radar sensor. With a standard computer, we achieve an update rate of about 90 Hz in our setup described below.

IV. Experiments

The above described method enables a simple, yet computationally very efficient radar-inertial state estimation with self-calibration capabilities (IMU intrinsics, radar-IMU extrinsics). In the following, we test our method on real platforms with real data.

A. Experimental setup

The sensor used for the experiments is the lightweight 60 GHz multi-channel FMCW radar transceiver Infineon BGT60, shown attached to the UAV in Fig. 1, mounted on the Infineon XMC4500 board equipped with a USB interface. The frequency spectrum of chirps generated by the radar is between $f_l = 57$ GHz and $f_u = 63$ GHz. We set the sampling frequency to $f_{sr} = 2$ kHz, the number of samples to $N_s = 200$ and the number of chirps to $N_c = 20$. For inertial measurements we use the IMU supplied on the PX4 platform. We set the sampling rate of the IMU to $f_{si} = 200$ Hz. For comparison with VIO, we use images grabbed by the onboard mvBlueFOX-MLC camera connected over USB. We set the frequency of the camera to $f_{sc} = 20$ Hz which is a reasonable choice from the potential on-board processing viewpoint. Although VIO can benefit from the stereo-camera setup, we do not make use of it because of the limited payload of the UAV. We set the camera exposure time to $e = 8\ ms$ for a typical indoor scene.

We placed three corner reflectors (Fig. 2) as radar targets at the locations $CR_1 = [x = 0.23, y = 0.65, z = 0.36]$m,

$CR_2 = [x = 0.52, y = 0.53, z = 0.48]$m and $CR_3 = [x = 0.48, y = 0.86, z = 0.59]$m.

We then acquired two datasets performing highly aggressive hand-held trajectories with the platform in Fig. 1. Each acquisition involved very sharp and aggressive movements (norm of max. angular velocity $\omega = 12.6\ \frac{rad}{s}$, norm of max linear acceleration $a = 33.0\ \frac{m}{s^2}$) of the sensor rig such that high motion blur was affecting the camera sensor. Fig. 4 depicts the setup from the onboard camera view with motion blur above the well textured area (left) and a sharp image on the low textured area (right). The trajectories were carried out just above the set of three corner reflectors (Fig. 5) allowing the same (visual) feature set to be visible during the entire experiment and thus allowing the VIO to leverage persistent features for locally non-drifting estimation. A sample range-Doppler reading by the radar of such a scene is depicted in Fig. 6. The two acquired datasets are recorded with feature-poor and feature-rich backgrounds in order to see the effect the background has on the VIO performance.

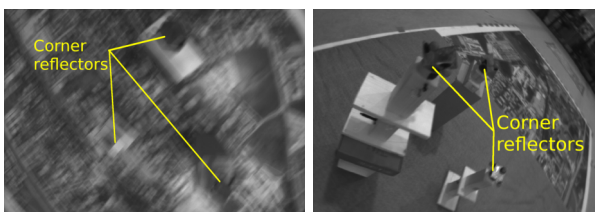

Fig. 4. Radar target setup shown from the onboard camera view. Left: above well textured area during a high-speed motion causing significant motion blur. Right: still phase above low textured area.

During acquisition, we recorded sensor readings from the camera, IMU, and radar together with the poses of the UAV and corner reflectors streamed by the optitrack system for ground truthing. Both estimators, our EKF-based radar-inertial state estimation and OpenVins [19] as state of the art reference VIO are run offline on the recorded sensor data on an Intel core i7 vPRO laptop with 15 GB RAM. For a fair comparison of the proposed methodology with the VIO, we carefully fine tuned the feature tracker (incl. persistent features) of OpenVins to get the best possible result out of the VIO algorithm given the high aggressiveness of the performed movements and the challenging amount of motion blur contained in the camera images.

B. Evaluation

First, we evaluate if our proposed data association yields adequate results to be used as update information in our EKF estimator framework. For this, we compare the measured ranges using the approach described in Sec. III-D to the ground truth range computed from the Optitrack measurements of the radar sensor and the corresponding target 3D position. Fig. 7 depicts in red the ground truth ranges during a 25 second long test and shows in blue the measurements obtained with our method. Note that often, our approach could not clearly associate the measured ranges to a target (gaps between the blue stars). This is due to the noise in the radar reading and our gating approach such

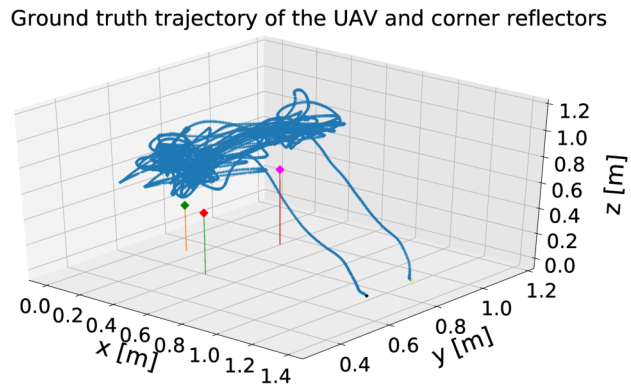

Fig. 5. Ground truth trajectory of the UAV above the set of three corner reflectors. Trajectory of the UAV is plotted in blue, green, magenta and red diamonds are the corner reflectors placed on the floor, the black and yellow dots are final and initial UAV positions respectively.

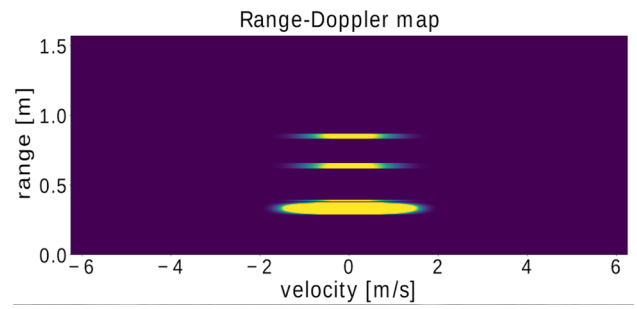

Fig. 6. Range-Doppler map generated from a typical scene as shown e.g. in Fig. 4 from the three corner reflectors. The ellipses in the velocity dimension are fairly wide which reflects high uncertainty of the Doppler velocity measurement. Thus, and because of the lighter computation, we only use range measurements in our approach.

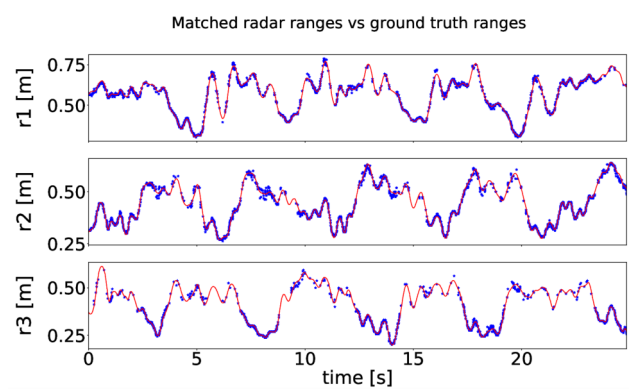

Fig. 7. Ranges measured with radar matched to the ground truth. Ground truth is plotted in red, matched radar range in blue.

that any measurement differing more than 3 cm to any possible target range is directly discarded (see Sec. III-D). That said, our tightly coupled estimator approach is resilient to intermittent target losses as any available measurement is seamlessly used as 1D measurement whenever available. Fig. 7 shows this particularly well around t=15 s where only the first target yields measurements. From this data we measured a standard deviation of $\sigma_m = 1\ cm$ for the measurements. This noise value was included in the EKF

TABLE I

EXPERIMENTS AND RMSE AFTER CONVERGENCE

	pos [m]	roll [deg]	pitch [deg]	yaw [deg]	drift [%]
well textured scene					
Ours	0.0268	0.7861	1.3476	1.2105	7.0985
VIO	0.0894	0.9272	0.6389	3.7205	13.2782
poorly textured scene					
Ours	0.0311	1.1901	1.3047	1.4743	5.1859
VIO	0.2183	1.1429	0.7438	8.9878	19.9222

process. As Fig. 6 shows, Doppler velocity could additionally be extracted from the radar signal. However, it significantly increased computational load for its calculation, as it requires to compute the FFT across all ranges for each chirp, and the fact that velocity is already observable with only range measurements are all factors that do not motivate to use the Doppler velocity as measurement. With our implementation we achieve an average range measurement rate of about 90 Hz.

Second, we test the estimator performance when the extracted ranges are fused together with IMU in an EKF framework. We compare this radar-inertial estimation with visual-inertial odometry for highly aggressive maneuvers. Fig. 8 and Fig. 9 show the absolute position error plots for our approach (red) and VIO (green) compared to the Optitrack ground truth in well-textured and poorly-textured scenes respectively. With the well tuned VIO, we managed to achieve non-diverging results for the VIO algorithm in both cases. A non-negligible drift (final error divided by overall path length), however, persists for both cases: about 13% for the well-textured scene, and nearly 20% for the poorly-textured one (see Tab. I for details). Fig. 10 shows a similar plot for the attitude error in the well-textured test. The unobservable yaw is most affected by the challenging data since the persistent features are lost due to the motion blur. We encourage to not take the figures and Tab. I as direct comparisons between the two algorithms since this would compare unobservable states in VIO against observable ones in our approach. Rather, they show the behavior of the approaches in challenging situations: for VIO they show an order of magnitude higher drift than usually reported in literature discussing well-behaving scenarios despite data-set specific tuning of the VIO algorithm. The increase in drift is caused by the VIO not being able to consistently keep the persistent features because the high motion blur. With the lack of their locally non-drifting information, the algorithm thus goes back to the mode in which only odometry information can be used. For our approach, the figures show an RMSE below 3 cm in position despite the very agile motion and only using sequential 1D range measurements for IMU integration correction.

Third, we evaluated the self-calibration capability of the proposed estimator. Fig. 11 shows the evolution of the extrinsic calibration state p_{IS} representing the 3D translation between the onboard IMU and radar sensor. After a wrong initialization, the state converges well.

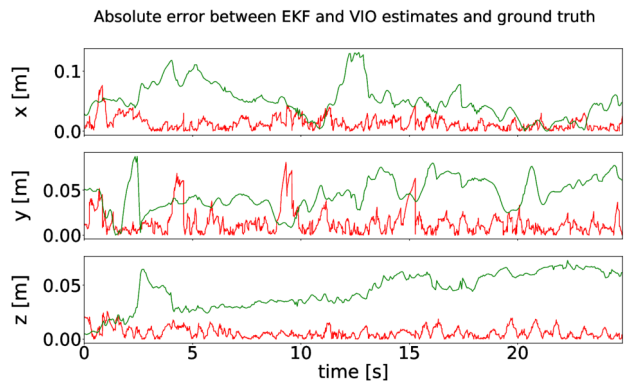

Fig. 8. Absolute position estimation error for feature-rich scenario. Errors for VIO are plotted in green, for our radar based approach in red. The position drift of the VIO is more than 13% showing the impact of the challenging data.

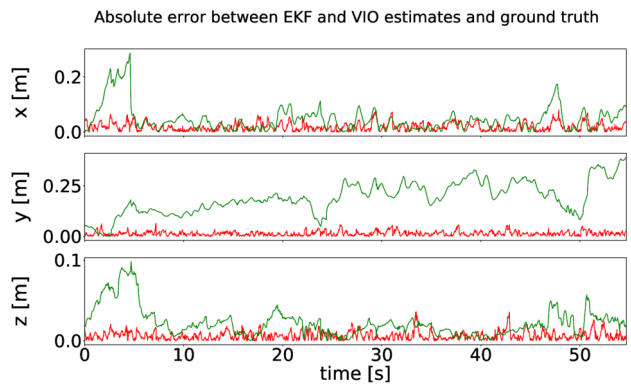

Fig. 9. Absolute position estimation error for feature-poor scenario. Errors for VIO are plotted in green, for our radar based approach in red. With our best tuning efforts, we managed to get non-diverging results and a position drift of about 20% for VIO.

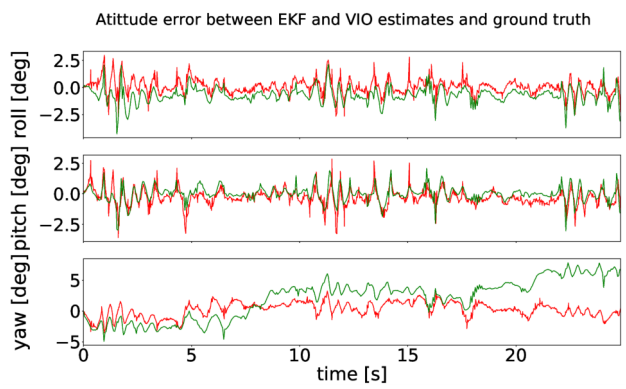

Fig. 10. Attitude estimation error for the feature-rich scenario. Errors for VIO are plotted in green, for our radar based approach in red. The challenging motion clearly affects the (unobservable) yaw drift of the VIO.

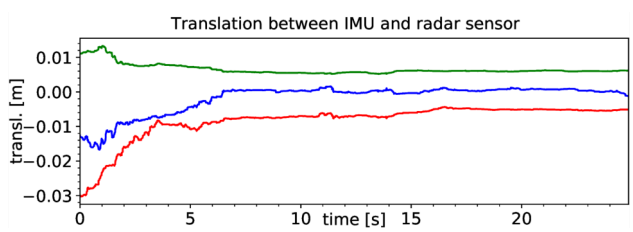

Fig. 11. Estimated translation p_{IS} between radar and IMU sensors for one of the experiments (feature-rich scenario). x, y and z coordinates are plotted in red, green and blue respectively.

V. Conclusions

In this paper, we presented a method which uses a tightly-coupled formulation of an Extended Kalman Filter in which we fuse measurements from an IMU with range measurements from a low-cost and lightweight FMCW radar. We presented a data association method to match targets in the environment with the signals in a radar scan. To eliminate any gauge freedom, we fixed at least two targets defining the gravity aligned world frame and can seamlessly include additional target positions due to the tightly coupled estimator formulation. We showed that this approach enables fast and accurate estimation of the UAV pose even in scenarios where VIO suffers from the image blur adversely affecting its accuracy (despite using persistent features for locally non-drifting information). Our computationally simple approach only extracting range information from the raw radar signals and use them as sequential 1D measurements in an EKF formulation enabled estimator update rates of about 90 Hz. Our EKF formulation is capable of estimating the navigation states, IMU intrinsics, and radar sensor extrinsics. The fast and motion-blur free measurements are particularly relevant for UAVs performing aggressive manoeuvres. As the next steps, moving from this rather area-bound proof of concept presented here, we will integrate radar sensors with longer range and perform larger tests that dynamically include multiple targets in an online estimation process to fully leverage the potential of the here presented approach.

Acknowledgment

The authors would like to thank Infineon Technologies AG for the support regarding the radar sensor and providing the BGT60 radar sensor.

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. 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[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] 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.

[11] Ş. Săftescu, M. Gadd, D. De Martini, D. Barnes, and P. Newman, "Kidnapped radar: Topological radar localisation using rotationally-invariant metric learning," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 4358–4364.

[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] 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.

[14] F. Uysal, P. Aubry, and A. Yarovoy, "Accurate target localization for automotive radar," 04 2019, pp. 1–5.

[15] 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.

[16] C. Doer, R. König, G. Trommer, and E. Stumpf, "Autonomous precision takeoff and landing system for vtols in degraded visual and gnss denied environments," 09 2020.

[17] F. She, Y. Zhang, D. Shi, H. Zhou, X. Ren, and T. Xu, "Enhanced relative localization based on persistent excitation for multi-uavs in gps-denied environments," *IEEE Access*, vol. PP, pp. 1–1, 08 2020.

[18] A. W. Doerry, "Reflectors for sar performance testing." Sandia National Laboratories, Tech. Rep., 2008.

[19] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "Openvins: A research platform for visual-inertial estimation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 4666–4672.

[20] C. Doer and G. F. Trommer, "Radar inertial odometry with online calibration," in *2020 European Navigation Conference (ENC)*, 2020, pp. 1–10.

[21] ——, "x-rio: Radar inertial odometry with multiple radar sensors and yaw aiding," vol. 12, 02 2022, pp. 329–339.

[22] A. Martinelli, "Visual-inertial structure from motion: observability and resolvability, iros 2013," *Tokyo, Japan*.

[23] J. Blueml, A. Fornasier, and S. Weiss, "Bias compensated uwb anchor initialization using information-theoretic supported triangulation points," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021.

[24] C. Stetco, B. Ubezio, S. Mühlbacher-Karrer, and H. Zangl, "Radar sensors in collaborative robotics: Fast simulation and experimental validation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 10 452–10 458.

[25] S. Weiss, M. W. Achtelik, M. Chli, and R. Siegwart, "Versatile distributed pose estimation and sensor self-calibration for an autonomous mav," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 31–38.

[26] 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.