

# Multi-Radar Inertial Odometry for 3D State Estimation using mmWave Imaging Radar

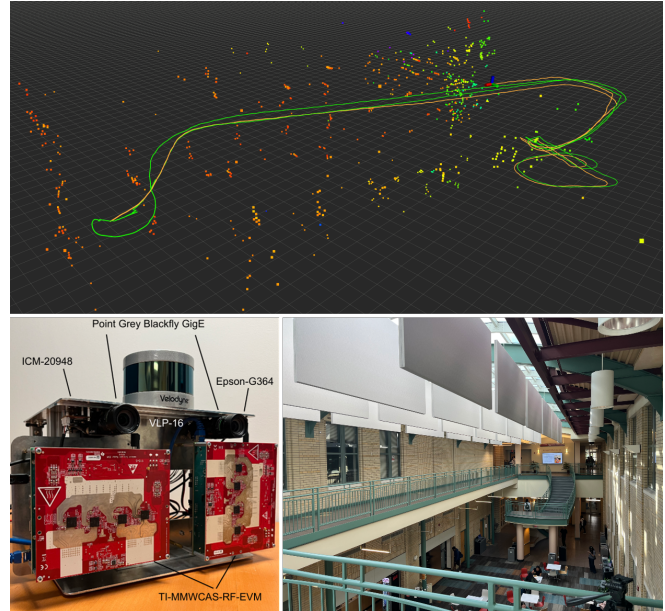
Jui-Te Huang, Ruoyang Xu, Akshay Hinduja, and Michael Kaess

**Abstract**—State estimation is a crucial component for the successful implementation of robotic systems, relying on sensors such as cameras, LiDAR, and IMUs. However, in real-world scenarios, the performance of these sensors is degraded by challenging environments, e.g. adverse weather conditions and low-light scenarios. The emerging 4D imaging radar technology is capable of providing robust perception in adverse conditions. Despite its potential, challenges remain for indoor settings where noisy radar data does not present clear geometric features. Moreover, disparities in radar data resolution and field of view (FOV) can lead to inaccurate measurements. While prior research has explored radar-inertial odometry based on Doppler velocity information, challenges remain for the estimation of 3D motion because of the discrepancy in the FOV and resolution of the radar sensor. In this paper, we address Doppler velocity measurement uncertainties. We present a method to optimize body frame velocity while managing Doppler velocity uncertainty. Based on our observations, we propose a dual imaging radar configuration to mitigate the challenge of discrepancy in radar data. To attain high-precision 3D state estimation, we introduce a strategy that seamlessly integrates radar data with a consumer-grade IMU sensor using fixed-lag smoothing optimization. Finally, we evaluate our approach using real-world 3D motion data.

## I. INTRODUCTION

State estimation serves as a fundamental component in the majority of robotics applications. Commonly used sensors for state estimation include cameras, LiDAR, and IMUs [1]. However, deploying robots can be challenging in adverse environments. For instance, LiDAR sensors can experience substantial performance degradation when exposed to adverse factors such as smoke and fog, as well as in environments lacking distinct geometric features. Similarly, cameras encounter the same challenges with additional difficulties in environments with low-light conditions or lacking distinct visual features.

The emerging 4D mmWave imaging radar sensor [2] can provide robust perception in demanding environments. The imaging radar sensor utilizes electromagnetic waves with wavelengths at the millimeter level that can function in adverse environmental conditions and do not rely on external lighting. Using frequency-modulated continuous wave technology, the imaging radar can provide Doppler velocity measurements for each detected 3D point. However, radar point clouds are known for being noisy and sparse and can be severely discretized due to post-processing procedures,



**Fig. 1:** A demonstration of multi-radar inertial odometry (orange) walked through different levels of an atrium compared to visual-inertial odometry (green). The colors of the radar point cloud indicate Doppler velocity from high (red) to low (purple).

thereby providing limited geometry information, especially in confined indoor environments.

Despite the challenges presented by the mmWave imaging radar, several recent studies have found success using this technology for object detection [3], [4], navigation [5] and state estimation [6]–[9]. Further applications are discussed in the survey paper [10].

There are several challenges in developing a state estimation algorithm using mmWave radar. Previous works demonstrated the capability of building trajectories based on geometry information and scan-to-scan matching in automotive settings [11]–[13]. However, these methods are limited to planar localization and often exhibit degraded performance in indoor environments where there are limited key point features and clear geometry patterns in addition to the noise introduced by multi-path reflections of the signals.

Another popular approach involves leveraging Doppler speed information to optimize body frame velocity and then fusing it with an inertial measurement sensor [6], [14], [15]. However, inaccuracies in body frame velocity and its associated uncertainty can lead to trajectory drift along the elevation direction. This issue arises because the antenna of imaging radar is designed to provide significantly lower data resolution and FOV in the elevation dimension compared to

This work was partially supported by Amazon Lab126.

J. Huang, A. Hinduja and M. Kaess are, R. Xu was, with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA {juiteh, ruoyangx, ahinduja, kaess}@cs.cmu.edu

the azimuth dimension [16].

To achieve high-precision 3D motion state estimation using a Doppler velocity-based radar-inertial odometry approach, we tackle the above-mentioned challenges. We analyze radar measurement uncertainty and develop an algorithm for seamless radar-IMU fusion. We showcase the effectiveness of this approach by using dual mmWave cascade imaging radars coupled with a consumer-grade IMU sensor to obtain accurate 3D state estimates. Our contributions can be summarized as follows:

- **Doppler and Velocity Uncertainty:** We assess the uncertainty of radar Doppler measurements and the resulting uncertainty on estimated linear body frame velocity for fusion with IMU sensor data.
- **Multi-Radar Inertial Odometry:** We employ a fixed-lag smoother optimization strategy capable of fusing IMU and multiple radar data to compensate for measurement uncertainty from each other and ensure robustness to outlier measurements.
- **Evaluate with 3D Motion:** We evaluate our method with hand-held collected data while traversing through different levels of indoor buildings.

## II. RELATED WORK

### A. Radar-Inertial Odometry

Doppler velocity information from mmWave radar point clouds has been used to estimate trajectory in previous research. Kramer et al. [14] proposed a sliding window optimization approach with a robust loss kernel to fuse IMU and radar Doppler velocity measurements. They demonstrate the estimated body frame velocity is close to a Visual-Inertial solution. Doer et al. [15] separately estimated the body frame velocity from radar points using a RANSAC method to remove outlier measurements introduced by noise. Then they proposed an EKF-based approach to fuse radar body frame velocity with IMU and barometer measurements to estimate drone poses in indoor environments. However, their results show difficulty on estimating height correctly without the assistance of barometer data.

Based on the previous EKF-based approach, Zhuang et al. further integrated the full radar SLAM system: 4D iRIOM [17]. The authors proposed using a Graduated Non-Convexity method to remove outlier Doppler velocity measurements and estimate velocity from radar. They also introduce scan-to-map matching using radar points in the system to improve trajectory estimation. The loop closure was done using Scan Context [18] to detect similar geometry on the radar points map. They demonstrate their methods on a small ground robot which is limited to a planar environment.

Recent research by Michalczyk et al. [19], [20] proposed to build 3D landmarks from radar point clouds and fuse them with IMU using multi-state EKF. However, this method required a controlled environment where persistent landmarks with high radar cross sections are required.

A learning-based approach [9] by Lu et al. was proposed to fuse IMU and radar measurements to estimate trajectory. The

authors use convolutional neural networks to build feature encoders for radar and recurrent neural networks to process IMU and the fused feature embeddings. However, their method was compared in [6] and found to have difficulty generalizing to new environments.

### B. Multi-Radar for State Estimation

To further improve state estimation using mmWave radar sensor. Adding multiple radar sensors in an attempt to increase the FOV has appeared in some research. Doer et al. based on their previous work EKF-RIO [15] proposed the x-RIO [7] using a triple radar setting to increase the horizontal FOV of the robot. The setting enables them to use the Manhattan World Assumption [21] on the radar point clouds to improve the trajectory estimation. In [22] Ng et al. use a sliding window optimization approach to fuse 4 radar measurements with an IMU to jointly estimate body frame velocity and odometry in an autonomous car application. However, both of the above multi-radar methods were applied only in the planar setting and did not address the elevation drift problem caused by relatively lower resolution on the elevation dimension.

Park et al. [6] proposed a solution to improve 3D ego-motion estimation by adding a second ground-facing radar that operates in synchronization with the primary radar. They only extract 2D velocity on the high-resolution azimuth dimension from each radar to fuse with IMU. They proposed a radar velocity factor that fuses radar velocity with the IMU gyroscope. The authors evaluated their method by traversing through different levels of a construction site to prove their performance with 3D motion. In comparison with their method, we present a solution for fusing multiple radars and IMU data without discarding high-frequency accelerometer measurements and without the need to synchronously trigger the radars.

The publicly available dataset for radar state estimation Coloradar dataset [8] provides a cascade imaging radar and a single-chip imaging radar both placed horizontally. The dataset was collected with mostly planar motion. Therefore, it's necessary to design our own radar sensors rig in order to collect 3D motion radar data for the research of radar-inertial state estimation.

## III. METHODOLOGY

In this section, we introduce our approach to estimating trajectories using a Radar-Inertial system. Firstly, for each radar in the system, we estimate the linear velocity in the sensor frame and its associated uncertainty using radar Doppler information. Subsequently, the estimated linear velocity is constructed as a keyframe to be fused with the IMU accelerometer and gyroscope data, serving as a measurement constraint in the integration process. The complete pose is integrated and optimized in a sliding window using the aforementioned measurements. The full states to be optimized in each radar keyframe are as follows:

$$\mathcal{X}_i = \left[ \mathbf{R}_i, \mathbf{p}_i, \mathbf{v}_i, \mathbf{b}_i^a, \mathbf{b}_i^g \right] \in \mathbb{R}^{15} \quad (1)$$

### A. Velocity Estimation using Radar Doppler Velocity

One special property of the Imaging Radar is the Doppler velocity of the detected points. The physical significance of the Doppler velocity is the projection of the sensor's linear velocity in the direction of the point. Assuming that most of the points in one frame are static objects, we can establish the relationship between the Doppler velocity  $\mathbf{d}_n \in \mathbb{R}^1$  of the point  $n$ , its 3D position  $\mathbf{r}_n \in \mathbb{R}^3$ , and the sensor's linear velocity  $\mathbf{v}_j^s \in \mathbb{R}^3$  in frame  $j$  with the following equation:

$$-\mathbf{d}_n = (\mathbf{v}_j^s)^\top \frac{\mathbf{r}_n}{\|\mathbf{r}_n\|} \quad (2)$$

To obtain an accurate linear velocity of the sensor that aligns with the most measured Doppler velocities, we can use a least-squares approach to solve for  $\mathbf{v}^s$ . The noise of Doppler measurement can be modeled by a Gaussian distribution with variance  $\Sigma_d \in \mathbb{R}^1$ . The optimized sensor velocity will be:

$$\mathbf{v}^s = \underset{\mathbf{v}^s}{\operatorname{argmin}} \left( \sum_n \|(\mathbf{v}^s)^\top \frac{\mathbf{r}_n}{\|\mathbf{r}_n\|} + \mathbf{d}_n\|_{\Sigma_d}^2 \right) \quad (3)$$

We can find the marginalized covariance  $\Sigma_{\mathbf{v}^s} \in \mathbb{R}^{3 \times 3}$  on variable  $\mathbf{v}^s$  by inverting the information matrix in the system.

$$\Sigma_{\mathbf{v}^s} = (A^\top \Sigma_d^{-1} A)^{-1} \quad (4)$$

Where  $A \in \mathbb{R}^{n \times 3}$  and each row of  $A$  is a unit vector directing to the radar point  $\mathbf{r}_n / \|\mathbf{r}_n\|$ . Notice that due to the radar antenna design, the radar points distribution is uneven in azimuth dimension and elevation dimension. As a result, the estimated velocity uncertainty exhibits variations along the XYZ axes. Further details regarding the evaluation of Doppler measurement uncertainty  $\Sigma_d$  and the estimated velocity uncertainty will be discussed in IV-B.

In reality, radar data often includes noisy measurements and non-stationary objects. To remove these outliers, previous works have explored the RANSAC approach [7], non-linear optimization with Cauchy robust loss [14], and a Graduated Non-Convexity (GNC) method [17]. While RANSAC is non-deterministic, GNC requires extra iterations to adjust the kernel, and the Cauchy robust loss is sensitive to the initial value setting. Considering computation efficiency and accuracy, we use a Cauchy robust loss kernel to remove outliers and use the Levenberg–Marquardt optimizer to solve the system. The initial value of the variable  $\mathbf{v}^s$  was set using the preintegrated body frame velocity from IMU measurements and rotated to the sensor's coordinate. Please refer to the next section for IMU integration.

Finally, to fuse the sensor velocity from different radars with IMU. We treat the IMU frame as the body frame. The linear velocity and the corresponding covariance on the body frame are as follows:

$$\mathbf{v}^b = \mathbf{R}_r \mathbf{v}^s \quad (5)$$

$$\Sigma_{\mathbf{v}^b} = \mathbf{R}_r \Sigma_{\mathbf{v}^s} \mathbf{R}_r^\top \quad (6)$$

Where  $\mathbf{R}_r \in SO(3)$  is the rotation from the radar sensor coordinate to the IMU coordinate.

### B. Preintegrated IMU Factor

As introduced in [23], we can efficiently fuse the IMU sensor with other low frame rate measurements using preintegration on manifold to avoid the relinearization procedure when the keyframes linearization point changes. With sensor measured linear acceleration  $\tilde{\mathbf{a}}_k$  and angular velocity  $\tilde{\omega}_k$  between keyframe  $i$  and the next keyframe  $j$ . The preintegrated measurement of relative position  $\Delta \tilde{\mathbf{p}}_{ij}$ , orientation  $\Delta \tilde{\mathbf{R}}_{ij}$ , and velocity  $\Delta \tilde{\mathbf{v}}_{ij}$  are:

$$\Delta \tilde{\mathbf{R}}_{ij} = \prod_{k=i}^{j-1} \operatorname{Exp}((\tilde{\omega}_k - \mathbf{b}_i^g) \Delta t) \quad (7)$$

$$\Delta \tilde{\mathbf{v}}_{ij} = \sum_{k=i}^{j-1} \Delta \tilde{\mathbf{R}}_{ik} (\tilde{\mathbf{a}}_k - \mathbf{b}_i^a) \Delta t \quad (8)$$

$$\Delta \tilde{\mathbf{p}}_{ij} = \sum_{k=i}^{j-1} \left( \Delta \tilde{\mathbf{v}}_{ik} \Delta t + \frac{1}{2} \tilde{\mathbf{R}}_{ik} (\tilde{\mathbf{a}}_k - \mathbf{b}_i^a) \Delta t^2 \right) \quad (9)$$

Where  $\mathbf{b}_i^a$  and  $\mathbf{b}_i^g$  are the slow varying linear acceleration bias and angular velocity bias. The measurement constraint of the IMU between 2 keyframes has the residual in the following form.

$$\mathbf{r}_{\Delta \mathbf{R}_{ij}} = \operatorname{Log} \left( \Delta \tilde{\mathbf{R}}_{ij}^\top (\mathbf{R}_i^\top \mathbf{R}_j) \right) \quad (10)$$

$$\mathbf{r}_{\Delta \mathbf{v}_{ij}} = \mathbf{R}_i^\top (\mathbf{v}_j - (\mathbf{v}_i + \mathbf{g} \Delta t_{ij})) - \Delta \tilde{\mathbf{v}}_{ij} \quad (11)$$

$$\mathbf{r}_{\Delta \mathbf{p}_{ij}} = \mathbf{R}_i^\top \left( \mathbf{p}_j - \left( \mathbf{p}_i + \mathbf{v} \Delta t_{ij} + \frac{1}{2} \mathbf{g} \Delta t_{ij}^2 \right) \right) - \Delta \tilde{\mathbf{p}}_{ij} \quad (12)$$

$$\mathbf{r}_{\Delta \mathbf{b}_{ij}^a} = \mathbf{b}_j^a - \mathbf{b}_i^a \quad (13)$$

$$\mathbf{r}_{\Delta \mathbf{b}_{ij}^g} = \mathbf{b}_j^g - \mathbf{b}_i^g \quad (14)$$

The combined preintegrated IMU residual is written as:

$$\mathbf{r}_{\Delta \mathbf{I}_{ij}} = \left[ \mathbf{r}_{\Delta \mathbf{R}_{ij}}, \mathbf{r}_{\Delta \mathbf{v}_{ij}}, \mathbf{r}_{\Delta \mathbf{p}_{ij}}, \mathbf{r}_{\Delta \mathbf{b}_{ij}^a}, \mathbf{r}_{\Delta \mathbf{b}_{ij}^g} \right] \in \mathbb{R}^{15} \quad (15)$$

With covariance  $\Sigma_{\Delta \mathbf{I}_{ij}} \in \mathbb{R}^{15 \times 15}$  which takes into account the noise in the estimated bias used for integration, it also preserves the correlation between the bias uncertainty and the preintegrated measurements' uncertainty. Part of the covariance matrix related to the IMU bias variables is used to describe slow-varying bias evolution, the magnitude of which is proportional to the preintegration time  $\Delta t_{ij}$ . More details about IMU preintegration on manifold can be found in [23].

### C. Body Frame Velocity Factor

Upon receiving any body frame velocity  $\mathbf{v}_j^b$  estimated from the radar, we build a keyframe out of the preintegrated IMU measurements. This body frame velocity constraint will be added between the integrated global velocity variable  $\mathbf{v}_j$  and the rotation part of the pose variable  $\mathbf{R}_j \in SO(3)$ . By the definition of body frame velocity, the residual of this factor at frame  $j$  is:

$$\mathbf{r}_{\mathbf{v}_j^b} = \mathbf{R}_j^\top \mathbf{v}_j - \mathbf{v}_j^b \quad (16)$$

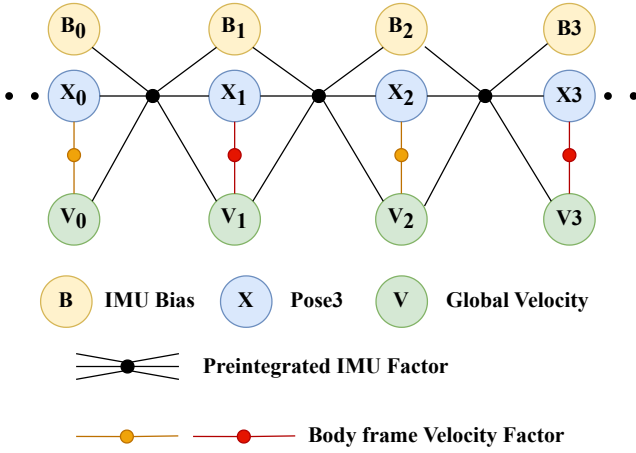


Fig. 2: An illustration of our multi-radar inertial state estimation system in the form of a factor graph. The body frame velocity factor can be built with either our horizontal or vertical imaging radar.

Occasionally, inaccurate estimates of the body frame velocity lead to outlier residuals. To counteract this issue, we incorporate a Huber loss kernel on this body frame velocity residual when solving the system.

To linearize the system, the Jacobian matrix with respect to rotation  $\mathbf{J}_R$  and global velocity  $\mathbf{J}_v$  are:

$$\mathbf{J}_R = [\mathbf{R}^\top \mathbf{v}]_\times \quad (17)$$

$$\mathbf{J}_v = \mathbf{R}^\top \quad (18)$$

Where  $[\mathbf{R}^\top \mathbf{v}]_\times$  is a skew symmetric matrix

$$[\mathbf{R}^\top \mathbf{v}]_\times = [\mathbf{q}]_\times = \begin{bmatrix} 0 & -q_z & q_y \\ q_z & 0 & -q_x \\ -q_y & q_x & 0 \end{bmatrix} \quad (19)$$

#### D. IMU Static Initialization

Since our preintegrated IMU factor is only constrained by the integrated velocity without a direct constraint within or among poses, an incorrect initial bias and rotation can easily cause the system to fail at the beginning. As IMU biases are modeled as slow-varying variables, it requires more steps to converge to the correct scale. Therefore, we have adapted and deployed a static initialization strategy for the IMU based on [24], utilizing solely IMU measurements.

The initialization process involves two consecutive sliding windows collecting IMU measurements. The second window, which includes the latest acceleration measurements, is used to detect any sensor movement by measuring the acceleration variance. Once motion is detected, we average the acceleration measurements in the first window to determine the gravity vector and its orthonormal basis using the Gram-Schmidt process. The initial rotation is determined by taking the inverse of the  $SO(3)$  rotation formed by the orthonormal basis. The initial accelerometer bias is calculated by rotating the gravity constant using the initial rotation and then subtracting it from the measured gravity vector. The initial gyroscope bias is simply the mean of the angular velocities in the first measurement window. We have found this initialization procedure to be particularly crucial for uncalibrated IMUs with larger biases.

#### E. Factor Graph Formulation

As IMU and radar measurements are inherently asynchronously triggered in time, IMU measurements were interpolated to be temporally aligned with radar measurements. This procedure guarantees that IMU measurements will always be available to construct a preintegrated IMU factor, even when two radar measurements are close and no IMU measurement exists in between.

Given all the estimated measurements and their covariance. The full system optimizes states in a sliding window  $\mathcal{S}$  and minimizes their Mahalanobis distance:

$$\mathcal{X}_S^* = \operatorname{argmin}_{\mathcal{X}_S} \left[ \sum_{i,j \in \mathcal{S}} \left( \|\mathbf{r}_{\Delta \mathbf{I}_{ij}}\|_{\Sigma_r}^2 + \|\mathbf{r}_{\mathbf{V}_j^b}\|_{\Sigma_{v^b}}^2 \right) \right] \quad (20)$$

Where body frame velocity covariance  $\Sigma_{v^b}$  was used for proper fusion with preintegrated IMU Factor.

An illustration of the system as a factor graph is shown in Fig. 2. The aforementioned system was constructed using the GTSAM library [25]. The optimization problem is solved by a fixed-lag smoother with iSAM2 [26]. We set a 5-second optimization window where variables passed the window will be marginalized as prior factors.

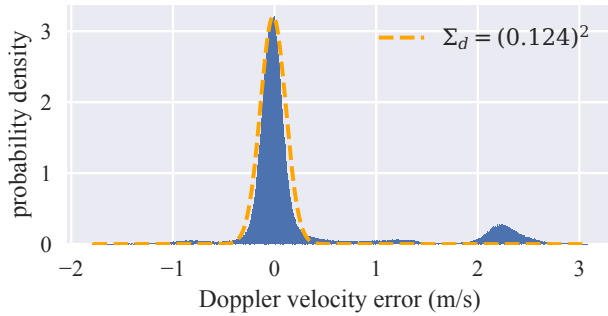
## IV. EXPERIMENTS

### A. Dual Cascade Imaging Radar Dataset

We designed our handheld sensor platform to collect time-synchronized radar, IMU, and image data for evaluating multi-radar inertial state estimation. The platform includes a Velodyne VLP-16 LiDAR, two PointGrey Cameras running at 20Hz, an Epson G-364 IMU running at 200Hz, an ICM-20948 IMU running at 100Hz, and two Texas Instrument Cascade Imaging Radars (MMWCAS-RF-EVM) running at 10Hz. The sensor platform layout is depicted in Fig. 1. Timing is synchronized using an external Teensy MCU with pseudo GPS/PPS signals and trigger signals for the cameras.

The commercially available TI mmWave cascade imaging radar has a theoretical azimuth/elevation angular resolution of  $1.4^\circ/18^\circ$  [16]. These resolutions represent the minimum angle between two equally large targets at the same range that the radar can distinguish and separate from each other. To compensate for this limitation, we employ two radars placed horizontally and vertically in this research. Our radar operates in Multi-Input Multi-Output (MIMO) mode. To avoid frequency conflicts, we set the start frequency of the horizontal and vertical radars as 77GHz and 79GHz, with both radars having a chirp slope of 40MHz/ $\mu$ s and a bandwidth of approximately 1.5GHz. With these radar signal settings, we achieve a maximum range of 30m, a Doppler velocity resolution of 0.055m/s, and a maximum Doppler velocity range of  $\pm 1.76$ m/s. We employed a Constant False Alarm Rate (CFAR) algorithm provided by the TI mmWave studio to process the raw radar data into point clouds, including signal-to-noise ratio and Doppler velocity information for each point.

Unlike previous work [6], where one radar was directed downward to the ground, we designed both radars to face



**Fig. 3:** The probability density of the Doppler velocity error distribution is calculated by comparing the velocities of 1,092,224 radar points with the VIO body-frame velocity projected in their respective directions.

forward. This decision was made to accommodate robots with limited space for installing downward-facing sensors. Additionally, our state estimation results suggest potential future enhancements in imaging radar designs to improve resolution in both azimuth and elevation dimensions.

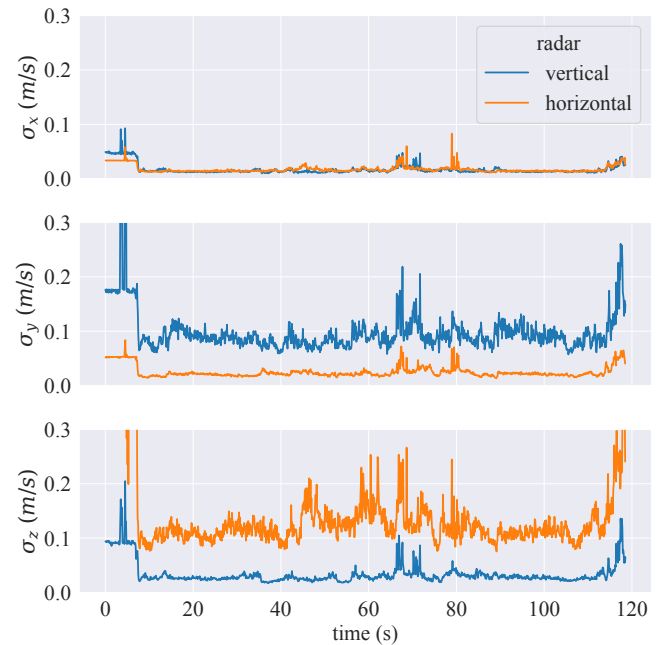
We collected three sequences with mostly planar motion and three sequences involving 3D motion across various levels within the building. The data collection was conducted at a normal human walking pace and motion. The traverse length of each trajectory is detailed in TABLE I.

### B. Doppler and Velocity Uncertainty Evaluation

To properly fuse all radar measurements with IMU measurements, we need to understand the noise levels of both sensors. The IMU measurement uncertainty can be found in manufacturers’ manuals or determined by analyzing hours of static IMU data [27]. However, current research and the radar sensor manufacturer do not provide Doppler velocity measurement uncertainty, and the Doppler velocity resolution does not accurately represent it.

To assess radar Doppler velocity errors, we utilize body-frame velocity data obtained from visual-inertial odometry pseudo ground truth. By projecting this velocity onto the direction of all radar points during a two-minute sequence (NSH\_atrium), we can analyze the Doppler velocity error distribution, as depicted in Fig. 3. The smaller peak in this bimodal distribution is the result of Doppler velocity values exceeding sensor limits. After excluding such outliers, we approximate the noise as a Gaussian distribution with variance  $\Sigma_d \approx (0.124 \text{ m/s})^2$ . It’s important to note that this uncertainty may vary with different radar signal settings, depending on factors such as Doppler velocity resolution and elevation/azimuth angle resolution.

We use this Doppler velocity noise model to compute the marginalized covariance on the optimized sensor frame velocity, and then the corresponding uncertainty on body frame velocity follows equations (4) and (6). Fig. 4 displays the standard deviation of the body frame linear velocity estimated from two different radars in a two-minute sequence. For each radar, the axis with high uncertainty aligns with the elevation direction. This emphasizes the importance of using multiple radars to compensate for inaccurate measurements from each other and to fuse them with the correct covariance scale. We also noticed that uncertainty increases when there



**Fig. 4:** Standard deviation on XYZ axes of the estimated body frame velocity from two radars.

are fewer radar measurements, such as at the sequence’s start and end when sensor motion is nearly stationary, and around the 70-second mark when the radar faces an open area for a short period.

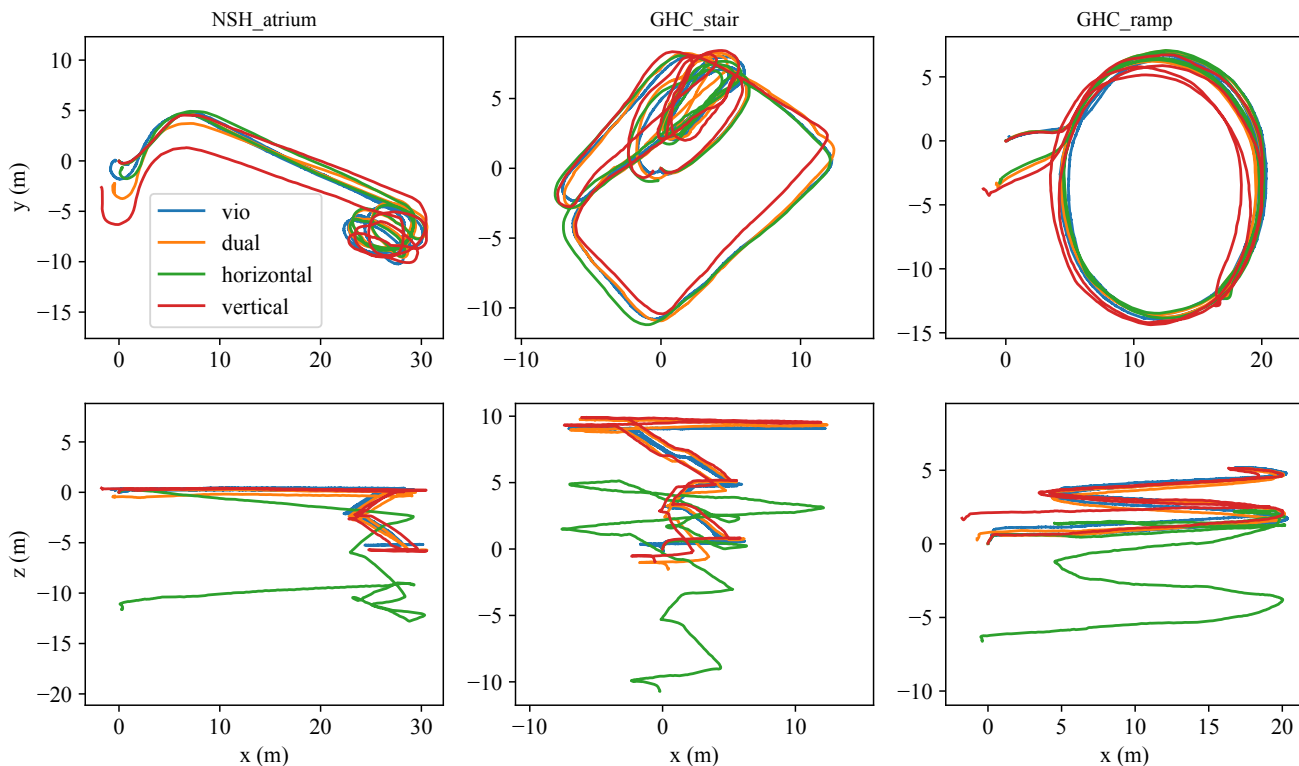
### C. Quantitative Evaluation

Given that our evaluation environment is indoors and includes long corridors, LiDAR point cloud registration becomes challenging in environments with such geometric features. Therefore we evaluate our method using pseudo ground truth generated by the stereo visual-inertial odometry system OpenVINS [24], with two PointGrey cameras and the Epson IMU. Our radar-inertial system is running with ICM-20948 IMU. The final trajectory was assessed using the EVO library [28], without the use of any alignment algorithms. The results for absolute and relative pose errors are presented in TABLE I. From the results, it’s apparent that using both radars leads to a slight decrease in rotation performance. However, we are able to significantly improve translational drift compared to using a single radar setting.

We had considered comparing our method with, to the best of our knowledge, the only open-sourced EKF-based radar-inertial odometry algorithm, X-RIO [7]. However, their algorithm was primarily designed for single-chip mmWave radar and requires handcrafted radar velocity uncertainty, such as adjustments for offset and maximum limit of the radar velocity covariance, defining the threshold for rejecting radar velocity updates. We chose not to include this comparison in our evaluation, as we were unable to generate meaningful results.

### D. Qualitative Evaluation

Fig. 5 displays the estimated trajectory of our method compared to visual-inertial odometry. From the top view, it’s clear that using only vertical radar results in increased translational drift compared to other settings, mainly due to



**Fig. 5:** Comparing the trajectories of visual-inertial odometry to radar-inertial odometry using either single horizontal/vertical radar or dual radars. Each column represents our 3D motion sequences, with the top row displaying the top-down view and the bottom row presenting the side view.

**TABLE I:** RMSE of APE and RPE in 6 different sequences

sequence & info	method	APE		RPE	
		trans	rot	trans	rot
2D FRC_1st 113 m	dual	<b>1.48</b>	2.73	<b>0.08</b>	0.66
	horizontal	3.46	<b>2.33</b>	0.15	<b>0.58</b>
	vertical	2.16	2.95	0.21	0.72
2D FRC_2nd 158 m	dual	<b>1.31</b>	3.76	<b>0.09</b>	0.65
	horizontal	2.88	<b>3.40</b>	0.12	0.69
	vertical	2.92	5.16	0.13	<b>0.63</b>
2D NSH_4th 167 m	dual	<b>1.29</b>	2.27	<b>0.08</b>	0.75
	horizontal	4.07	<b>1.52</b>	0.12	<b>0.66</b>
	vertical	1.77	2.74	0.12	0.68
3D NSH_atrium 139 m	dual	<b>1.09</b>	4.92	<b>0.10</b>	0.89
	horizontal	7.22	<b>2.33</b>	0.18	0.88
	vertical	1.93	5.38	0.22	<b>0.78</b>
3D GHC_stair 146 m	dual	<b>1.16</b>	8.13	<b>0.11</b>	0.63
	horizontal	6.25	<b>5.02</b>	0.14	<b>0.57</b>
	vertical	1.58	7.50	0.16	0.71
3D GHC_ramp 219 m	dual	<b>1.36</b>	7.23	<b>0.06</b>	<b>0.53</b>
	horizontal	4.09	<b>6.43</b>	0.09	0.60
	vertical	1.78	7.30	0.08	<b>0.53</b>

The units for APE: translation / rotation are meter and degree. The units for RPE: translation / rotation are percentage and degree per meter.

its limited capacity to estimate lateral velocity. On the other hand, employing only a horizontal radar yields excellent performance in the XY plane but leads to substantial drift along the z-axis due to its reduced capacity to estimate vertical velocity effectively.

Given that human walking motion involves minimal lateral movement, the introduction of more vertical motion when traversing different levels of a building can result in significant drift when exclusively using horizontally placed radar sensors. This highlights the importance of considering motion primitives when deciding where to install radars for Doppler velocity-based radar-inertial odometry.

The attached video shows mapping results using estimated poses and radar points that have Doppler velocity projection error under the threshold. In the video, the geometry of the environment emerges even with noisy radar point clouds.

## V. CONCLUSION

This paper presents a system using dual mmWave cascade imaging radars fused with an IMU sensor capable of achieving high-precision 3D motion state estimation. We provide insights into the limitations of radar measurements, emphasizing the resulting estimated uncertainty and the imperative need to compensate for inaccuracies through the use of multiple radars. We present our radar configurations and fixed-lag optimization solutions, which effectively integrate the radar and IMU measurements. Our method is demonstrated using a real-world 3D motion dataset. For further improvement, efforts can be made to explore the utilization of noisy radar point cloud geometry and to address issues related to point cloud registration degeneracy.

## REFERENCES

- [1] S. Zhao, H. Zhang, P. Wang, L. Nogueira, and S. Scherer, "Super odometry: IMU-centric LiDAR-visual-inertial estimator for challenging environments," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Prague, CZ, Sep. 2021, pp. 8729–8736.
- [2] C. Iovescu and S. Rao, "The fundamentals of millimeter wave sensors," *Texas Instruments*, pp. 1–8, 2017.
- [3] Y. Wang, Z. Jiang, X. Gao, J.-N. Hwang, G. Xing, and H. Liu, "RODNet: Radar object detection using cross-modal supervision," in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, USA, Jan. 2021, pp. 504–513.
- [4] Y. Wang, J.-N. Hwang, G. Wang, H. Liu, K.-J. Kim, H.-M. Hsu, J. Cai, H. Zhang, Z. Jiang, and R. Gu, "ROD2021 challenge: A summary for radar object detection challenge for autonomous driving applications," in *Proc. ACM Intl. Conf. on Multimedia Retrieval (ICMR)*, Taipei, TW, Nov. 2021, pp. 553–559.
- [5] J.-T. Huang, C.-L. Lu, P.-K. Chang, C.-I. Huang, C.-C. Hsu, P.-J. Huang, H.-C. Wang *et al.*, "Cross-modal contrastive learning of representations for navigation using lightweight, low-cost millimeter wave radar for adverse environmental conditions," *IEEE Robotics and Automation Letters (RA-L)*, vol. 6, no. 2, pp. 3333–3340, 2021.
- [6] Y. S. Park, Y.-S. Shin, J. Kim, and A. Kim, "3D ego-motion estimation using low-cost mmWave radars via radar velocity factor for pose-graph SLAM," *IEEE Robotics and Automation Letters (RA-L)*, vol. 6, no. 4, pp. 7691–7698, 2021.
- [7] C. Doer and G. F. Trommer, "x-RIO: Radar inertial odometry with multiple radar sensors and yaw aiding," *Gyroscope and Navigation*, vol. 12, pp. 329–339, 02 2022.
- [8] A. Kramer, K. Harlow, C. Williams, and C. Heckman, "ColoRadar: The direct 3d millimeter wave radar dataset," *Intl. J. of Robotics Research (IJRR)*, vol. 41, no. 4, pp. 351–360, 2022.
- [9] C. X. Lu, M. R. U. Saputra, P. Zhao, Y. Almalioglu, P. P. De Gusmao, C. Chen, K. Sun, N. Trigoni, and A. Markham, "milliEgo: single-chip mmWave radar aided egomotion estimation via deep sensor fusion," in *Proc. ACM Conf. on Embedded Networked Sensor Systems*, Yokohama, JP, Nov. 2020, pp. 109–122.
- [10] A. Venon, Y. Dupuis, P. Vasseur, and P. Merriaux, "Millimeter wave FMCW radars for perception, recognition and localization in automotive applications: A survey," *IEEE Trans. on Intelligent Vehicles (TIV)*, vol. 7, no. 3, pp. 533–555, 2022.
- [11] P.-C. Kung, C.-C. Wang, and W.-C. Lin, "A normal distribution transform-based radar odometry designed for scanning and automotive radars," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, Xi'an, CN, May 2021, pp. 14 417–14 423.
- [12] Z. Hong, Y. Petillot, and S. Wang, "RadarSLAM: Radar based large-scale SLAM in all weathers," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Las Vegas, USA, Sep. 2020, pp. 5164–5170.
- [13] S. H. Cen and P. Newman, "Radar-only ego-motion estimation in difficult settings via graph matching," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, Montreal, CA, 2019, pp. 298–304.
- [14] A. Kramer, C. Stahoviak, A. Santamaria-Navarro, A.-A. Agha-Mohammadi, and C. Heckman, "Radar-inertial ego-velocity estimation for visually degraded environments," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, Paris, FR, May 2020, pp. 5739–5746.
- [15] C. Doer and G. F. Trommer, "An EKF based approach to radar inertial odometry," in *Proc. Intl. Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Karlsruhe, DE, Sep. 2020, pp. 152–159.
- [16] I. Texas Instruments, "Imaging radar using cascaded mmwave sensor reference design," 2020.
- [17] Y. Zhuang, B. Wang, J. Huai, and M. Li, "4D iRIOM: 4D imaging radar inertial odometry and mapping," *IEEE Robotics and Automation Letters (RA-L)*, vol. 8, no. 6, pp. 3246–3253, 2023.
- [18] G. Kim and A. Kim, "Scan context: Egocentric spatial descriptor for place recognition within 3D point cloud map," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Madrid, ES, Oct. 2018, pp. 4802–4809.
- [19] J. Michalczyk, R. Jung, and S. Weiss, "Tightly-coupled EKF-based radar-inertial odometry," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Kyoto, JP, Oct. 2022, pp. 12 336–12 343.
- [20] J. Michalczyk, R. Jung, C. Brommer, and S. Weiss, "Multi-state tightly-coupled EKF-based radar-inertial odometry with persistent landmarks," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, London, UK, May 2023, pp. 4011–4017.
- [21] J. Coughlan and A. L. Yuille, "The Manhattan world assumption: Regularities in scene statistics which enable Bayesian inference," *Advances in Neural Information Processing Systems*, vol. 13, 2000.
- [22] Y. Z. Ng, B. Choi, R. Tan, and L. Heng, "Continuous-time radar-inertial odometry for automotive radars," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Prague, CZ, Sep. 2021, pp. 323–330.
- [23] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual-inertial odometry," *IEEE Trans. on Robotics (TRO)*, vol. 33, no. 1, pp. 1–21, 2016.
- [24] P. Geneva, K. Eickenhoff, W. Lee, Y. Yang, and G. Huang, "OpenVINS: A research platform for visual-inertial estimation," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, Paris, FR, May 2020, pp. 4666–4672.
- [25] F. Dellaert and GTSAM Contributors, "borglab/gtsam," May 2022. [Online]. Available: <https://github.com/borglab/gtsam>
- [26] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the bayes tree," *Intl. J. of Robotics Research (IJRR)*, vol. 31, no. 2, pp. 216–235, 2012.
- [27] O. J. Woodman, "An introduction to inertial navigation," University of Cambridge, Computer Laboratory, Tech. Rep., 2007.
- [28] M. Grupp, "evo: Python package for the evaluation of odometry and SLAM," <https://github.com/MichaelGrupp/evo>, 2017.