2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

RaCo-SLAM: A Physics-Informed 4D Radar SLAM with Co-visibility Consistency Factor

Zishun Deng[1], Wanbiao Lin[2], Can Li[1], Teng Wang[3], Chao Guo[4], Jiawei Shen[1], and Lei Sun[1]

***Abstract*— Robust all-weather localization is a critical capability for autonomous systems. While 4D mmWave radar offers superior resilience to adverse environmental conditions compared to LiDAR and cameras, its application in high-precision Simultaneous Localization and Mapping (SLAM) is hindered by significant challenges, including severe point cloud sparsity, complex noise characteristics, and the prevalence of dynamic objects. To address these issues, we propose RaCo-SLAM, a robust and real-time 4D mmWave radar SLAM framework with co-visibility consistency. This framework features a novel physics-informed probabilistic model for adaptive feature extraction from sparse and noisy point clouds. For global consistency, we introduce a co-visibility consistency factor (CoVC factor) into the global optimization, moving beyond conventional loop-closure methods. This factor directly minimizes point-to-point registration errors to enforce global consistency and is designed for parallel real-time execution on a standard CPU. Comprehensive evaluation on diverse and challenging real-world datasets demonstrates state-of-the-art accuracy and robustness, achieving real-time performance exceeding 40 Hz on a standard CPU. To benefit the community, the code and collected dataset will be released at https://github.com/sudo-robot0/RaCo-SLAM.**

I. Introduction

SLAM is a foundational technology for autonomous systems. While significant progress has been made with traditional sensors like cameras and LiDAR, their performance is notoriously susceptible to degradation in adverse conditions such as rain, snow, fog, or dust [1], [2]. This inherent limitation poses a critical bottleneck for applications requiring robust, all-weather autonomy.

Millimeter-wave (mmWave) radar, and in particular the emerging 4D imaging radar, has recently gained significant traction as a compelling alternative [3]. Operating at millimeter-level wavelengths, it is largely immune to the aforementioned environmental degradations. Moreover, modern 4D radars produce point clouds that are not only three-dimensional but also endowed with per-point radial velocity measurements via the Doppler effect. This intrinsic motion information is especially valuable for distinguishing static structures from dynamic objects and for stabilizing ego-motion estimation under low-feature conditions [4].

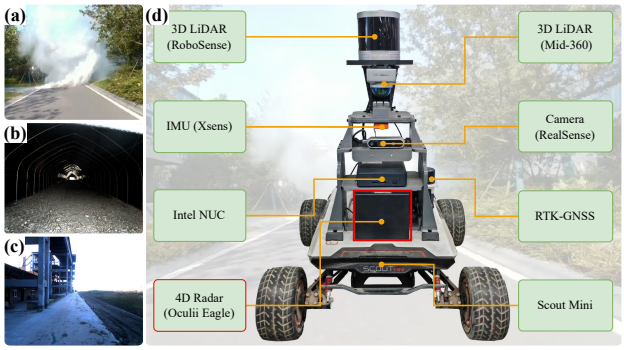


Fig. 1: Multi-sensor experimental platform and data collection scenarios. The ground robot (d) is equipped with a 4D mmWave radar, supplemented by a full sensor suite for ground truth acquisition. (a)-(c) illustrate the challenging and perceptually-degraded environments used for evaluation.

Despite these advantages, high-precision SLAM with 4D radar remains challenging. The generated point clouds are extremely sparse and contaminated by complex non-Gaussian noise, including multi-path reflections and ghost points. Furthermore, the typically narrow Field of View (FOV) aggravates the difficulty of reliable data association, leading to rapid accumulation of odometry errors. These factors hinder the extraction of stable geometric features—a cornerstone of many state-of-the-art SLAM algorithms—and exacerbate long-term drift. While several pioneering radar SLAM systems [5]–[9] have made progress, they still struggle with two persistent issues: *robust feature selection under high uncertainty* and *maintaining long-term global consistency*.

To address these challenges, we present RaCo-SLAM, a novel and robust 4D mmWave radar SLAM framework, developed and validated on our experimental platform shown in Fig. 1. Our approach comprises two main stages: a motion estimation module and a global optimization module. The motion estimation stage introduces a probabilistic likelihood model to intelligently select high-quality, static features by fusing Doppler velocity, radar cross-section (RCS), and intensity information. The global optimization stage departs from traditional pose graph methods by incorporating our novel co-visibility consistency factor, which directly minimizes point-to-point registration errors on the factor graph, thereby achieving global consistency without relying on explicit loop closures.

The key contributions are summarized as follows:

This work was supported by the National Natural Science Foundation of China (No. 62173192 and No. 62373365).

[1]Zishun Deng, Can Li, Jiawei Shen, and Lei Sun are with the Institute of Robotics and Automatic Information System, Nankai University, Tianjin 300350, China (e-mail: {2120240552, canli, 2120240564}@mail.nankai.edu.cn, sunl@nankai.edu.cn). *Corresponding author: Lei Sun.*

[2]Wanbiao Lin is with the Shenzhen Research Institute, Nankai University, Shenzhen, China (e-mail: 2120160381@mail.nankai.edu.cn).

[3]Teng Wang is with the Department of Automation, Tsinghua University, Beijing 100086, China (e-mail: tengwang2002@gmail.com).

[4]Chao Guo is with JiangHuai Advanced Technology Center, Hefei 230000, China (e-mail: guoc171@163.com).

- The development of RaCo-SLAM, a complete and real-time system equipped with a novel physics-informed probabilistic model for adaptive feature extraction from sparse and noisy point clouds. This design effectively accommodates drastic variations in point cloud density across different sensors and challenging environments.
- A co-visibility consistency factor for global optimization is proposed, which directly minimizes point-to-point registration errors to enable continuous drift correction and is designed for parallel real-time execution on a CPU.
- Comprehensive evaluation on diverse and challenging real-world datasets, including a newly collected dataset. Experimental results demonstrate the accuracy and robustness of the proposed system, and evaluations across multiple sequences provide insights into its performance and stability.

II. Related Work

A. Radar Odometry

Modern LiDAR odometry has achieved remarkable precision through two dominant paradigms: geometry-based methods that extract explicit features like planes and edges [10], [11], and direct methods that register raw point clouds [12]–[14]. Both paradigms capitalize on the dense, well-structured nature of LiDAR data, a luxury not afforded to 4D radar systems.

However, 4D radar odometry encounters substantial challenges due to the intrinsic sparsity and high noise of its point clouds. Early works adapted established techniques. For instance, several methods enhance the conventional Iterative Closest Point (ICP) algorithm by incorporating Doppler velocity measurements to better constrain scan registration [15], [16], while others modify ICP variants such as Generalized-ICP (GICP) to explicitly model measurement uncertainty [5]. Although a few recent approaches have begun to exploit RCS measurements to reject spurious correspondences [17], the majority of these methods fail to fully exploit the rich, physics-based information inherent in 4D radar scans. Instead, they often rely on simple, manually-tuned thresholding of Doppler or RCS values. This heuristic-based approach lacks adaptability across different sensor types and environments, and consequently struggles to distinguish truly static features from transient noise. This leaves the core challenges of feature uncertainty and robust data association largely unresolved. To address this gap, we propose a physics-informed probabilistic motion estimation module that adaptively incorporates these physical cues for robust and accurate odometry, regardless of the input point cloud's density or quality.

B. Globally Consistent Optimization

A global optimization module is essential to mitigate the inevitable long-term drift from the odometry. The dominant paradigm in SLAM, including for radar systems [5], is pose graph optimization constrained by explicit loop closures. However, this approach faces fundamental challenges in the context of 4D radar. First, place recognition for 4D radar is still an immature field; while some methods show promise [18]–[20], they often suffer from a high rate of false positives and negatives, especially in complex and dynamic environments. Second, many real-world trajectories do not contain loops, or loops are too sparse to effectively constrain drift. This leads to a significant delay in correction, often causing large, abrupt jumps in the trajectory when a loop is finally closed. Finally, compressing a rich point cloud match into a single relative pose with a Gaussian noise model represents a significant loss of information. An incorrect loop closure can catastrophically distort the entire map and trajectory.

To achieve higher accuracy, some works explore methods beyond simple pose graphs, such as adapting Bundle Adjustment to LiDAR. LiDAR Bundle Adjustment [21]–[23], while being the gold standard for accuracy, is computationally prohibitive and relies on the stable, long-term tracking of geometric features—a condition rarely met by sparse radar data. A more relevant concept in LiDAR is multi-view registration [24], which was effectively used in the odometry module of CFEAR [25] for 2D spinning radar odometry to improve local accuracy. However, this strategy is fundamentally a local one. It refines the pose of the current frame against a recent map but does not alter or correct the poses of historical keyframes in the trajectory. Thus, it cannot mitigate long-term, accumulated drift globally.

To the best of our knowledge, leveraging multi-view geometric consistency as a global constraint for 4D radar SLAM remains unexplored. In this work, we propose a co-visibility consistency factor, a novel factor that directly minimizes point-to-point errors between co-visible, non-consecutive keyframes on the factor graph. This provides a continuous and implicit geometric check, retaining maximal information while avoiding the fragility of explicit place recognition, thereby helping to ensure long-term global consistency.

III. Methodology

Our proposed SLAM framework, RaCo-SLAM, illustrated in Fig. 2, transforms sparse 4D radar scans into a globally consistent trajectory and map. The pipeline consists of a motion estimation stage for feature extraction (Sec. III-B) and local motion estimation (Sec. III-C), and a global optimization stage for trajectory refinement (Sec. III-D). The motion estimation stage first preprocesses each scan to extract stable static features, which are then fed to a scan-to-map registration module. The global optimization stage integrates the resulting odometry as a prior and our co-visibility consistency factors into a factor graph to correct long-term drift.

A. Problem Formulation

Let the state of the robot at time k be denoted by $\mathbf{x}_k$, which represents its pose $\mathbf{T}_k = [\mathbf{R}_k|\mathbf{t}_k] \in SE(3)$, where $\mathbf{R}_k \in SO(3)$ is the orientation and $\mathbf{t}_k \in \mathbb{R}^3$ is the position. The complete trajectory is the sequence of states $\mathbf{X} = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N\}$. The goal of our SLAM system is

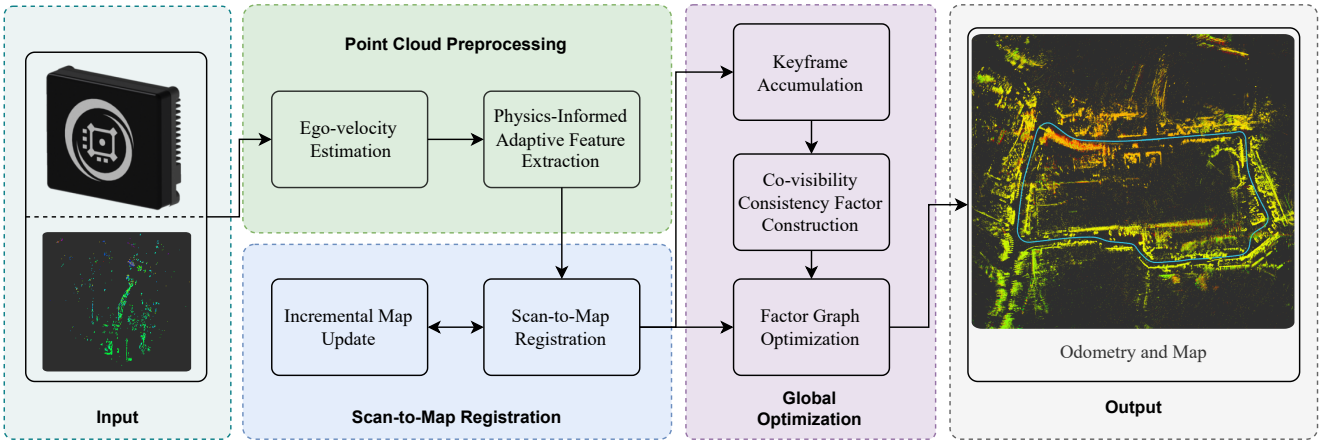


Fig. 2: System architecture of RaCo-SLAM. The system transforms sparse, noisy radar scans into a globally consistent map and trajectory via three main modules: (1) a point cloud preprocessing stage for ego-velocity estimation and probabilistic feature extraction, (2) a scan-to-map registration module for local motion estimation, and (3) a global optimization module, where a factor graph leverages our novel co-visibility consistency factors to ensure global consistency.

to find the maximum a posteriori (MAP) estimate of the trajectory $\mathbf{X}^*$ given all measurements, which is equivalent to solving a non-linear least-squares problem:

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \left(\sum_k \|e_{\text{odom}}(\mathbf{x}_{k-1}, \mathbf{x}_k)\|^2_{\mathbf{\Omega}_k} + \sum_{(i,j)\in\mathcal{C}} \|e_{\text{covis}}(\mathbf{x}_i, \mathbf{x}_j)\|^2_{\mathbf{\Lambda}_{i,j}} \right), \quad (1)$$

where e_{odom} is the error of an **odometry factor** with information matrix $\mathbf{\Omega}_k$, and e_{covis} is the error of a **co-visibility consistency factor**, weighted by its corresponding information matrix $\mathbf{\Lambda}_{i,j}$, between two overlapping keyframe poses $\mathbf{x}_i$ and $\mathbf{x}_j$ from the set of all co-visible pairs $\mathcal{C}$. Details of the co-visibility factor are provided in Sec. III-D.

B. Point Cloud Preprocessing

Reliable feature extraction is critical for radar SLAM due to the sparse and noisy nature of the point clouds. Furthermore, the point cloud density distributions vary significantly across different 4D mmWave radar sensors and environments. Our preprocessing pipeline is designed to be adaptive, distilling raw radar data into a sparse set of stable, static features.

1) Ego-Velocity Estimation from Doppler Measurements: The Doppler velocity v_d measured by a radar sensor for a point $\mathbf{p}$ is the projection of the relative velocity between the sensor and the point onto its line-of-sight vector:

$$v_d = \hat{\mathbf{u}}^T(\mathbf{v}_{\text{target}} - \mathbf{v}_{\text{ego}}), \quad (2)$$

where $\mathbf{v}_{\text{ego}} \in \mathbb{R}^3$ is the sensor's ego-velocity, $\mathbf{v}_{\text{target}} \in \mathbb{R}^3$ is the target point's true velocity, and $\hat{\mathbf{u}}$ is the unit direction vector to the point. By invoking the static world assumption ($\mathbf{v}_{\text{target}} = \mathbf{0}$), this simplifies to $-v_d \approx \hat{\mathbf{u}}^T\mathbf{v}_{\text{ego}}$. We formulate an overdetermined linear system from all static points in a scan and embed its least-squares solver within a RANSAC framework to robustly estimate $\hat{\mathbf{v}}_{\text{ego}}$. The resulting velocity estimate and the per-point residuals,

$$\epsilon_{v,j} = |(-v_{d,j}) - (\hat{\mathbf{u}}_j^T\hat{\mathbf{v}}_{\text{ego}})|, \quad (3)$$

are crucial inputs for the subsequent feature extraction stage.

2) Physics-Informed Adaptive Feature Extraction: With an ego-velocity, we can now select high-quality features via a two-stage pipeline: an adaptive voxelization and outlier rejection, followed by a probabilistic scoring stage for final feature selection.

a) Adaptive Voxelization and Outlier Rejection: To ensure a sufficient number of features across varying environments, we dynamically adjust the voxel size. The voxel size for the current frame k, $s_{v,k}$, is updated based on the size used in the previous frame, $s_{v,k-1}$, and the number of features it yielded, N_{k-1}, targeting a constant number N_{target}:

$$s_{v,k} = \text{Clamp}\left(s_{v,k-1} \cdot \left(\frac{N_{k-1}}{2N_{\text{target}}} + 0.5\right), s_{\min}, s_{\max}\right). \quad (4)$$

Here, $\text{Clamp}(x, a, b) = \min(\max(x, a), b)$ constrains the voxel size within $[s_{\min}, s_{\max}]$, preventing extreme values. This allows the system to adapt its selectivity: in sparse scenes ($N_{k-1} < N_{\text{target}}$), the voxel size decreases for a finer search, while in dense scenes ($N_{k-1} > N_{\text{target}}$), it increases to merge regions and extract only the most salient features.

After voxelization, we discard a voxel if the number of its occupied neighboring voxels within a $3 \times 3 \times 3$ grid falls below a threshold. This step efficiently removes isolated, spurious measurements.

b) Probabilistic Feature Likelihood Estimation: We propose a probabilistic framework to estimate the likelihood of a point $\mathbf{p}_j$ being a stable, static feature, denoted by the event S_j. The likelihood is based on a set of physical observations $\mathcal{O}_j$ associated with the point, including its velocity error residual $\epsilon_{v,j}$, its RCS c_j, and its intensity i_j. Assuming conditional independence, the overall likelihood $L(S_j|\mathcal{O}_j)$ is:

$$L(S_j|\mathcal{O}_j) = P(\epsilon_{v,j}|S_j) \cdot P(c_j|S_j) \cdot P(i_j|S_j). \quad (5)$$

Motion Likelihood. The term $P(\epsilon_{v,j}|S_j)$ models motion stability. A true static feature is expected to have a velocity error residual $\epsilon_{v,j}$ close to zero. We model this with a

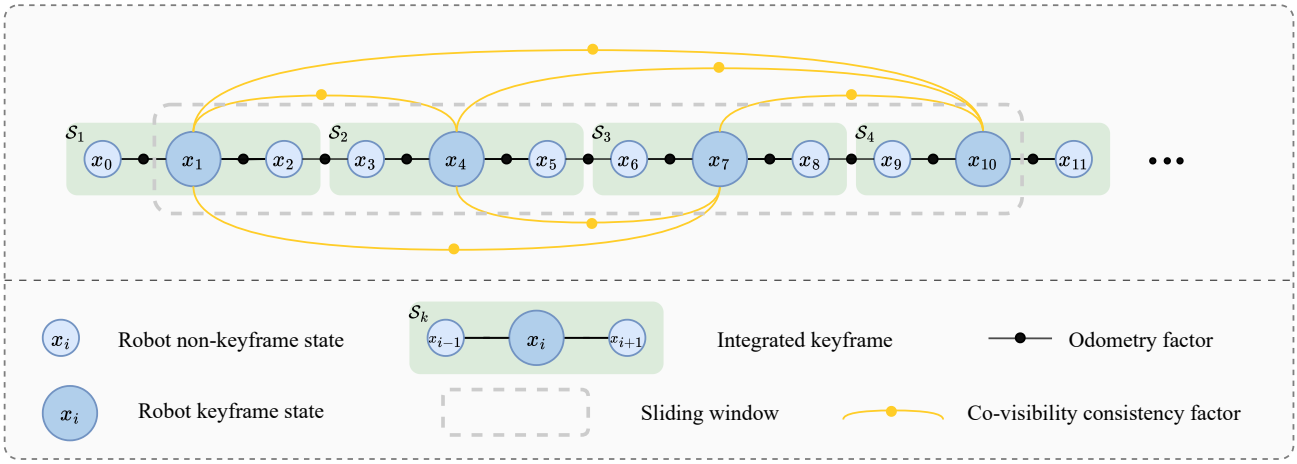


Fig. 3: Factor graph of RaCo-SLAM. A fixed number of consecutive frames (green) are first aggregated into a keyframe $\mathcal{S}_k$. Each keyframe corresponds to a keyframe state $\mathbf{x}_i$ (large deep blue node) in the factor graph. These keyframe states, along with intermediate non-keyframe states (small light blue nodes), are connected by odometry factors (black lines) that provide local motion estimates. To correct long-term drift, co-visibility consistency factors (yellow lines) are introduced. As illustrated for the current keyframe, e.g., $\mathbf{x}_{10}$, these factors connect it to co-visible historical keyframes (e.g., $\mathbf{x}_7, \mathbf{x}_4, \mathbf{x}_1$) within a sliding window (gray area), ensuring global consistency.

Laplacian distribution:

$$P(\epsilon_{v,j}|S_j) = \exp(-\alpha|\epsilon_{v,j}|), \tag{6}$$

where α is a positive scaling factor.

Physical Attribute Likelihoods. High-quality, persistent features (e.g., poles, guardrails, building facades) tend to yield high RCS and intensity returns [26]. We model the likelihood of observing RCS c_j and intensity i_j using sigmoid functions, which map these "higher is better" attributes to a score in $(0, 1)$:

$$P(c_j|S_j) = \sigma(\beta(c_j - c_{\text{mid},V})) = \frac{1}{1+\exp(-\beta(c_j - c_{\text{mid},V}))}, \tag{7}$$

$$P(i_j|S_j) = \sigma(\gamma(i_j - i_{\text{mid},V})) = \frac{1}{1+\exp(-\gamma(i_j - i_{\text{mid},V}))}, \tag{8}$$

where, for each non-empty voxel V, the terms $c_{\text{mid},V}$ and $i_{\text{mid},V}$ are the median of all points contained within that voxel. This allows the feature scoring to adapt to local variations in reflectivity. The parameters β and γ are global hyper-parameters that control the steepness of the likelihood curves.

For each non-empty, non-outlier voxel V, we compute the likelihood $L(S_j|\mathcal{O}_j)$ for every point $\mathbf{p}_j$ within it and select all points whose likelihood exceeds a confidence threshold:

$$\mathcal{F}_v = \{\mathbf{p}_j \in V \mid L(S_j|\mathcal{O}_j) > 0.3\}, \tag{9}$$

where $\mathcal{F}_v$ is the set of selected features from the voxel. The final feature set for the current scan is the union of feature sets from all voxels, $\mathcal{F} = \bigcup_v \mathcal{F}_v$.

C. Scan-to-Map Registration

To overcome point cloud sparsity, we register the current feature set $\mathcal{F}_k$ against a dynamic hash-based voxel map $\mathcal{M}_{k-1}$, which is aggregated from recent keyframes. For each feature point $\mathbf{p}_i \in \mathcal{F}_k$, its correspondence $\mathbf{q}_i \in \mathcal{M}_{k-1}$ is found via a nearest neighbor search. We then solve the following weighted, robust minimization problem to find the optimal transformation $\mathbf{T}_k^*$:

$$\mathbf{T}_k^* = \underset{\mathbf{T}_k \in SE(3)}{\arg\min} \sum_{(\mathbf{p}_i, \mathbf{q}_i)} w_i \cdot \rho_\delta \left(\|\mathbf{q}_i - \mathbf{T}_k \mathbf{p}_i\| \right), \tag{10}$$

where $\rho_\delta(\cdot)$ is the Huber robust loss function. Critically, the weight w_i is assigned the point's static feature likelihood $L(S_i|\mathcal{O}_i)$ from Eq. (5). This ensures that more reliable features, as determined by our probabilistic model, contribute more significantly to the optimization. The resulting transformation is used to construct an odometry factor for the global optimization.

D. Global Consistency Mapping

While the scan-to-map odometry provides accurate local motion estimates, long-term drift is inevitable. To ensure global consistency, we introduce a global optimization based on a factor graph, as illustrated in Fig. 3. Keyframe states in the graph correspond to integrated point clouds accumulated from several consecutive frames.

Crucially, our optimization departs from traditional methods that use pre-computed relative pose factors. Instead, we directly minimize the point cloud registration error through our novel co-visibility consistency factor, which connects co-visible keyframes within a sliding window and robustly constrains the global trajectory.

1) Keyframe Accumulation and Representation: To reduce the complexity of the factor graph and create denser point clouds for reliable matching, we accumulate high-likelihood feature points from a fixed-size window of consecutive frames into a single integrated keyframe. Specifically, the pose of the central frame within this window is designated as the pose for the integrated keyframe. Feature points from all other frames in the window are then transformed into the coordinate system of this central frame before being aggregated. Each integrated keyframe corresponds to a single state node in our factor graph. These aggregated point clouds are stored in a hash-based voxel map. To

maximize performance on multi-core processors, both the nearest-neighbor search and the hash map operations are parallelized using Intel Threading Building Blocks (TBB).

2) Co-visibility Consistency Factor: To construct our global consistency constraints, we first identify suitable co-visible keyframes. Our strategy operates within a fixed-size sliding window of historical keyframes to ensure bounded complexity. For the current keyframe $\mathcal{S}_j$, we search within this window for candidates $\mathcal{S}_i$ whose Euclidean distance is less than twice the sensor's effective range. Each potential pair then undergoes a geometric verification step by computing the number of well-matched point correspondences. If this number exceeds a threshold, the pair is confirmed as co-visible. For each confirmed pair $(\mathcal{S}_i, \mathcal{S}_j)$, we then introduce the **co-visibility consistency factor**. This factor directly minimizes the point-to-point registration error between the two keyframes, $\mathcal{S}_i$ and $\mathcal{S}_j$, within the factor graph. Unlike a standard `BetweenFactor` in GTSAM, our factor re-evaluates the error from the raw point clouds at every optimization step, thereby avoiding information loss.

The factor constrains the two absolute poses $\mathbf{x}_i$ and $\mathbf{x}_j$ by minimizing a cost function with a robust kernel $\rho(\cdot)$:

$$f_{\text{robust}}(\mathbf{x}_i, \mathbf{x}_j) = \sum_k \rho\left(\|\mathbf{d}_k(\mathbf{x}_i, \mathbf{x}_j)\|^2\right), \tag{11}$$

where $\mathbf{d}_k = \mathbf{p}'_k - \mathbf{x}_i^{-1}\mathbf{x}_j\mathbf{p}_k$ is the residual for the k-th correspondence between a point $\mathbf{p}_k \in \mathcal{S}_j$ and its corresponding point $\mathbf{p}'_k \in \mathcal{S}_i$. This is optimized using Iteratively Reweighted Least Squares (IRLS), which in each iteration constructs and solves a local quadratic approximation $f(\Delta\boldsymbol{\xi}) \approx C + \mathbf{b}^T\Delta\boldsymbol{\xi} + \frac{1}{2}\Delta\boldsymbol{\xi}^T\mathbf{H}\Delta\boldsymbol{\xi}$. The components of this quadratic form, the weighted block gradient vector $\mathbf{b}$ and the weighted block Hessian matrix $\mathbf{H}$, are computed using weights w_k derived from the robust kernel as:

- **Weighted Block Gradient vector** $\mathbf{b}$:

$$\mathbf{b} = \begin{bmatrix}\mathbf{b}_i \\ \mathbf{b}_j\end{bmatrix} = \sum_k w_k \begin{bmatrix}\mathbf{J}_{k,i}^T\mathbf{e}_k \\ \mathbf{J}_{k,j}^T\mathbf{e}_k\end{bmatrix}. \tag{12}$$

- **Weighted Block Hessian matrix** $\mathbf{H}$: Using the Gauss-Newton approximation, the Hessian is approximated as:

$$\begin{aligned}\mathbf{H} &\approx \sum_k w_k \begin{bmatrix}\mathbf{J}_{k,i}^T \\ \mathbf{J}_{k,j}^T\end{bmatrix}\begin{bmatrix}\mathbf{J}_{k,i} & \mathbf{J}_{k,j}\end{bmatrix} \\ &= \sum_k w_k \begin{bmatrix}\mathbf{J}_{k,i}^T\mathbf{J}_{k,i} & \mathbf{J}_{k,i}^T\mathbf{J}_{k,j} \\ \mathbf{J}_{k,j}^T\mathbf{J}_{k,i} & \mathbf{J}_{k,j}^T\mathbf{J}_{k,j}\end{bmatrix}.\end{aligned} \tag{13}$$

where $\mathbf{e}_k$ is the residual and $\mathbf{J}_{k,i}, \mathbf{J}_{k,j}$ are the Jacobians at the current linearization point with respect to perturbations on $\mathbf{x}_i$ and $\mathbf{x}_j$. These components are used to construct a `gtsam::HessianFactor` constraining the poses $\mathbf{x}_i$ and $\mathbf{x}_j$.

In our implementation, we employ the Cauchy kernel, where $\rho(x)$ is defined as:

$$\rho_{\text{cauchy}}(x) = \frac{c^2}{2}\log\left(1 + \frac{x^2}{c^2}\right), \tag{14}$$

with c being a tuning parameter. The corresponding IRLS weight is $w_k = \frac{c^2}{c^2 + \|\mathbf{e}_k\|^2}$, which enables the optimization to dynamically down-weight the influence of outlier point pairs, thereby yielding a more accurate and stable global pose estimate.

IV. Experiments

In this section, we conduct comprehensive experiments to rigorously evaluate the performance of our proposed RaCo-SLAM framework.

A. Experimental Setup and Datasets

Our evaluation is performed on three challenging public datasets and one self-collected dataset to ensure a comprehensive assessment. These include the NTU4DRadLM Dataset [27], designed for SLAM tasks with the Oculii Eagle radar; the MSC-RAD4R Dataset [28], providing diverse automotive scenarios also with an Oculii Eagle radar; and the HeRCULES Dataset [29], a multi-session benchmark featuring a Continental ARS548 4D radar. Furthermore, we collected our own challenging dataset in a mixed urban environment. As shown in Fig. 1, our platform is a ground robot equipped with an Oculii Eagle radar, two 3D LiDAR sensors, an IMU, a camera, and an RTK-GNSS receiver for ground truth.

All performance evaluations and algorithm runtimes were conducted on a laptop with an Intel Core i9-14900HX CPU and 32 GB RAM. Our algorithm was configured to use a maximum of 8 threads and no GPU acceleration, demonstrating its efficiency on a standard CPU.

B. Comparison with State-of-the-art Methods

We compare RaCo-SLAM against a suite of state-of-the-art methods. We include the work of Zhang et al. [5], evaluating both its odometry-only pipeline, hereafter referred to as **4DRadarOdom**, and its full SLAM implementation with loop closures, **4DRadarSLAM**. We also compare against **EKF-RIO** [4], a classic EKF-based radar-inertial odometry, and **EFEAR-4D** [8], a method featuring robust Doppler-based ego-velocity filtering and region-wise geometric feature extraction. This comparison is particularly insightful as it contrasts our continuous consistency approach using the co-visibility consistency factor against both pure odometry and traditional loop-closure-based optimization. Finally, we include **KISS-ICP** [14], a leading LiDAR odometry approach, which we apply directly to the radar data as a strong cross-modal baseline.

1) Quantitative Analysis: We present a comprehensive quantitative evaluation using the Absolute Pose Error (APE) metric [30]. The detailed results across all four datasets are presented in Table I, II, III, and IV. For clarity, 4DRadarOdom in the tables denotes the odometry-only pipeline from [5], while 4DRadarSLAM refers to the same system with its loop closure functionality enabled. Best results are in **bold** and second-best are underlined.

Across all major public datasets, RaCo-SLAM consistently achieves state-of-the-art performance. On the

NTU4DRadLM dataset (Table I) and our self-collected dataset (Table II), it is significant that our approach, without any explicit loop closure module, outperforms the full 4DRadarSLAM system. Notably, on the particularly challenging `loop2` sequence, which covers a 4.23 km trajectory, our method achieves a final drift rate of 1.14% (48.153 m APE), demonstrating superior long-term drift correction capabilities compared to all baselines. This directly validates our core hypothesis, outlined in Sec. I and II, that achieving global consistency through continuous geometric constraints from co-visible keyframes is more robust and effective than relying on sparse and potentially unreliable place recognition events.

TABLE I: RMSE OF ABSOLUTE POSE ERROR (APE) [m] ON THE NTU4DRADLM DATASET

Method	cp	garden	loop2	Mean
4DRadarOdom	2.167	2.311	131.948	45.475
4DRadarSLAM	0.702	**1.112**	50.616	17.477
EKF-RIO	6.241	6.032	436.262	149.512
EFEAR-4D	1.546	1.292	77.131	26.656
KISS-ICP	2.609	2.452	82.996	29.352
RaCo-SLAM	**0.608**	1.133	**48.153**	**16.631**

TABLE II: RMSE OF ABSOLUTE POSE ERROR (APE) [m] ON OUR SELF-COLLECTED DATASET

Method	sequence 01	sequence 02	sequence 03	Mean
4DRadarOdom	0.495	2.211	3.299	2.002
4DRadarSLAM	0.482	1.355	3.301	1.713
EKF-RIO	2.624	8.932	15.907	9.154
EFEAR-4D	2.512	2.293	2.928	2.578
KISS-ICP	0.871	1.369	2.075	1.438
RaCo-SLAM	**0.195**	**0.814**	**1.129**	**0.713**

The results on the more diverse and dynamic MSC-RAD4R dataset (Table III) further highlight our robustness. While EFEAR-4D shows strong performance in short, snowy rural sequences (e.g., `RURAL_C2`–`RURAL_F2`), this is largely because the short trajectory length limits drift. In contrast, our system demonstrates superior consistency across a wider variety of scenarios. In particular, it achieves robustness in the challenging `SMOKE_A0` and `SMOKE_B0` sequences, which feature heavy smoke that reduces point cloud density. This is attributed to our physics-informed probabilistic model, which adaptively selects stable features even with numerous moving objects and in perceptually degraded environments.

The most compelling results are on the challenging HeRCULES dataset (Table IV). RaCo-SLAM dominates the comparison, achieving the best APE on every sequence. The primary challenge here is the severe sparsity of the Continental ARS548 radar, which provides only around 250 points per frame, substantially fewer than the 6000+ points from the Oculii Eagle radar. This sparsity makes robust geometric feature extraction nearly impossible, explaining the poor performance of most radar-based methods. Notably, KISS-ICP performs reasonably well, as its simple point-to-point registration is less dependent on rich geometric structures. However, RaCo-SLAM still outperforms it decisively because our physics-informed model leverages Doppler and RCS data to extract reliable features even from extreme sparsity, while our global optimization ensures superior global consistency.

2) Qualitative Analysis: The limitations of traditional loop-closure-based methods are qualitatively illustrated in Fig. 4. In the NTU4DRadLM `cp` sequence (Fig. 4a), 4DRadarSLAM correctly detects a loop but introduces an inconsistent constraint that distorts the map, while our method maintains a more accurate trajectory. In the MSC-RAD4R `SMOKE_A0` sequence (Fig. 4b), a clear loop exists but is missed by the traditional place recognition module due to perceptual degradation from the smoke, causing 4DRadarSLAM to accumulate significant drift. Our system, which does not depend on explicit loop detection, remains consistent. Finally, in corridor-like trajectories without loops (Fig. 4c), loop-closure methods have no opportunity to correct drift, while our RaCo-SLAM effectively constrains errors throughout the entire sequence using its continuous co-visibility consistency factors.

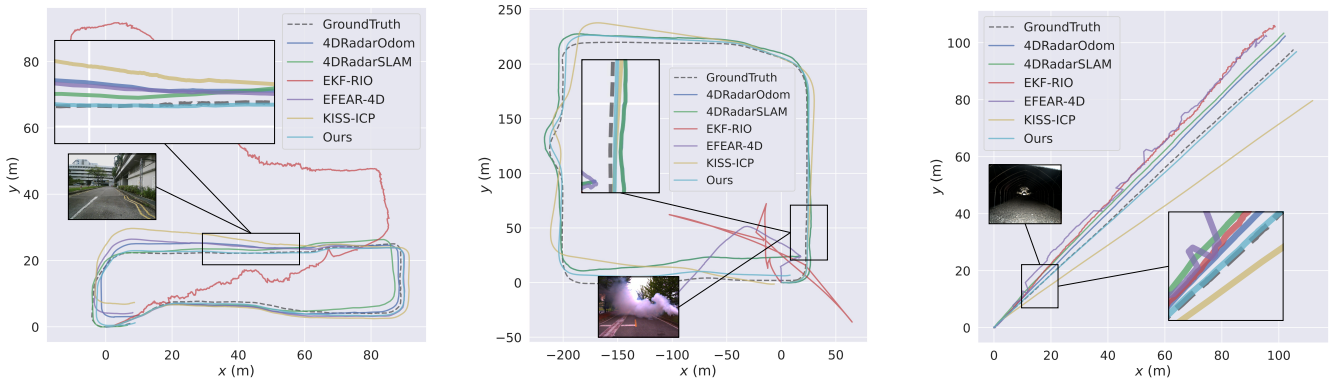
The final mapping quality serves as a tangible indicator of the system's global consistency. As shown in Fig. 5a, our optimization module constructs numerous co-visibility consistency factor constraints between co-visible keyframes, forming a robust geometric backbone. The resulting high-quality map, when overlaid on a satellite image (Fig. 5b), shows excellent alignment with real-world structures, further confirming the high global accuracy of our framework.

C. Ablation Study

To validate the effectiveness of our core contributions, we conduct a thorough ablation study on four representative sequences. We evaluate several variants of our system against the full RaCo-SLAM framework:

- **w/o Adaptive-Preprocessing**: This variant disables our probabilistic model and adaptive voxelization, replacing them with a simple Doppler filter and a fixed voxel size.
- **w/o CoVC factor (Odometry-only)**: We disable our co-visibility consistency factors in the global optimization, relying only on the odometry.
- **RaCo-SLAM (using BetweenFactor)**: This variant uses a standard `gtsam::BetweenFactor` as the co-visibility constraint, which computes a relative pose only once and adds it as a fixed constraint, to test the benefit of our continuous error re-evaluation.

The results, presented in Table V and visualized in Fig. 6, clearly demonstrate the impact of each component. Disabling our adaptive model ('w/o Adaptive-Preprocessing') degrades performance across all sequences, confirming that our probabilistic model and adaptive voxelization are critical for robustly extracting high-quality features. The most significant performance drop is observed in the 'w/o CoVC factor (Odometry-only)' variant, highlighting that our global optimization module is essential for correcting long-term drift. For instance, the APE on the `cp` sequence increases by nearly 300% from 0.608 m to 2.408 m without the optimization. Most importantly, the full RaCo-SLAM outperforms the 'RaCo-SLAM (using BetweenFactor)' variant. This is



(a) NTU4DRadLM `cp` **(b)** MSC-RAD4R `SMOKE_A0` **(c)** Ours `Sequence 01`

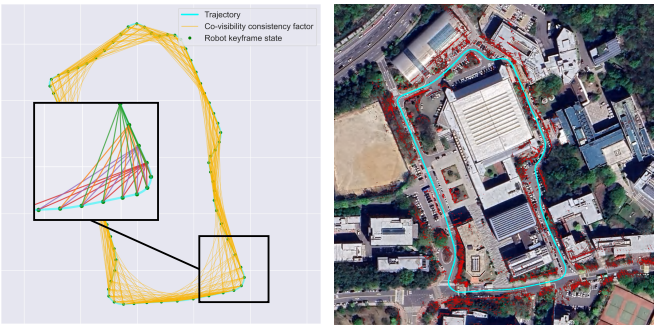
Fig. 4: Qualitative comparison of global consistency on three challenging sequences. Our method demonstrates superior robustness and accuracy by leveraging continuous co-visibility constraints in all cases: (a) 4DRadarSLAM detects a loop but introduces inconsistencies, while our method maintains higher accuracy. (b) A loop exists but is missed by the traditional method, leading to significant uncorrected drift. (c) In a trajectory with no loops, loop-closure-based methods cannot correct drift, while our RaCo-SLAM effectively constrains errors throughout.

TABLE III: RMSE OF ABSOLUTE POSE ERROR (APE) [m] ON THE MSC-RAD4R DATASET

Method	SMOKE_A0	SMOKE_B0	RURAL_A2	RURAL_B2	RURAL_C2	RURAL_D2	RURAL_E2	RURAL_F2	URBAN_A0	LOOP_D0	Mean
4DRadarOdom	7.485	5.745	37.961	45.575	1.785	2.215	1.064	0.431	8.388	22.465	13.311
4DRadarSLAM	7.323	5.481	17.435	49.186	2.013	2.496	1.037	0.452	7.976	28.380	12.178
EKF-RIO	94.245	87.536	278.788	251.471	11.245	8.537	7.745	7.045	18.638	26.895	79.215
EFEAR-4D	100.353	91.248	17.960	16.389	**0.404**	0.335	**0.275**	**0.298**	**2.428**	**10.973**	24.066
KISS-ICP	7.149	7.717	11.627	18.688	0.441	0.555	0.483	0.439	3.426	11.151	6.168
RaCo-SLAM	**2.968**	**3.418**	**11.536**	**14.720**	0.432	**0.325**	0.288	0.314	2.541	13.644	**5.019**

TABLE IV: RMSE OF ABSOLUTE POSE ERROR (APE) [m] ON THE HERCULES DATASET

Method	Library 01	Library 02	Library 03	Parking Lot 01	Parking Lot 02	Parking Lot 03	Parking Lot 04	Sports Complex 01	Sports Complex 02	Sports Complex 03	Mean
4DRadarOdom	76.239	62.474	21.608	22.888	22.245	20.199	19.346	85.357	34.723	70.125	43.520
4DRadarSLAM	67.155	54.655	19.131	22.842	21.537	18.789	19.255	83.833	32.402	69.725	40.932
EKF-RIO	78.042	76.254	74.788	29.999	27.862	31.657	25.193	97.440	77.216	84.740	60.319
EFEAR-4D	94.164	108.369	71.520	25.576	25.036	25.568	18.739	91.332	73.777	74.297	60.838
KISS-ICP	11.094	3.695	7.146	1.133	0.702	1.131	1.555	5.831	1.938	4.798	3.902
RaCo-SLAM	**4.877**	**2.882**	**3.094**	**0.783**	**0.550**	**1.074**	**1.006**	**4.619**	**1.527**	**2.974**	**2.339**



(a) CoVC factor connections in the factor graph **(b)** Map overlay on satellite image

Fig. 5: Global optimization and final mapping result for the HERCULES `Sports Complex 02` sequence.

TABLE V: ABLATION STUDY: RMSE OF APE [m]

Method Variant	cp	SMOKE_A0	Library 02	Sequence 01
RaCo-SLAM (Full)	**0.608**	**2.968**	**2.882**	**0.195**
w/o Adaptive-Preprocessing	0.775	3.312	3.527	0.264
w/o CoVC factor (Odometry-only)	2.408	4.779	4.469	0.524
RaCo-SLAM (using BetweenFactor)	1.738	4.001	3.892	0.754

a crucial finding, as it proves the benefit of our factor's continuous error re-evaluation. By compressing the rich point cloud match into a single, fixed relative pose constraint, the 'BetweenFactor' version loses significant geometric information. In contrast, our full system retains this information and continuously re-evaluates the registration error, leading to a more accurate and globally consistent trajectory.

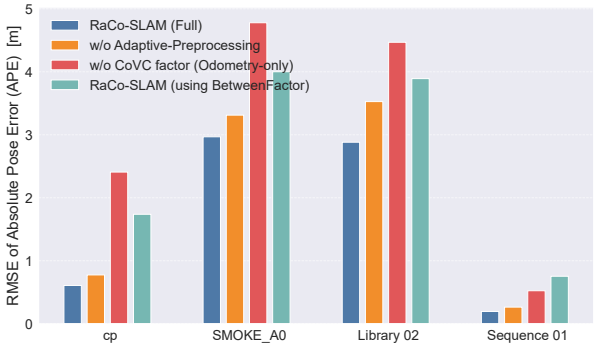


Fig. 6: Ablation study results. The full RaCo-SLAM framework consistently achieves the lowest error, while removing the CoVC factor ('w/o CoVC factor (Odometry-only)') results in the most significant performance degradation.

D. Real-time Performance

We analyze the runtime performance of RaCo-SLAM on the computational platform described in Section IV-A. Table VI shows the average processing time per frame for each main module across several representative sequences from different datasets and sensors. The entire system achieves a

real-time rate between 43 Hz and 123 Hz, demonstrating its suitability for real-world applications. Notably, processing the sparse Continental ARS548 data is significantly faster. The parallelization of our global optimization using TBB was critical to achieving this stable and efficient performance.

TABLE VI: AVERAGE RUNTIME PER FRAME [ms]

Module	**cp** (Oculii, 15Hz)	**LOOP_D0** (Oculii, 15Hz)	**Library 01** (Continental, 20Hz)	**sequence 01** (Oculii, 15Hz)
Preprocessing	1.23	1.27	0.20	1.21
Registration	10.14	9.70	2.34	10.97
Global Optimization	11.65	10.13	5.55	11.06
Total (Avg./Frame)	**23.02**	**21.10**	**8.09**	**23.24**
Frequency (Hz)	43.45	47.38	123.67	43.04

V. Conclusion

In this paper, we introduced RaCo-SLAM, a novel, complete, and real-time SLAM framework for 4D mmWave radar. Our system addresses the core challenges of radar-based localization through a physics-informed probabilistic motion estimation module for robust feature extraction and a novel co-visibility consistency factor for continuous global drift correction. By directly minimizing point cloud registration errors on the factor graph, our method achieves superior global consistency compared to traditional approaches that rely on sparse and fragile loop closures. Comprehensive experiments on multiple large-scale public datasets and our own collected data demonstrated that RaCo-SLAM consistently outperforms existing state-of-the-art radar SLAM systems while running in real-time on a standard CPU. Future work will explore the extension of the co-visibility concept to multi-robot SLAM.

References

[1] L. Fan, J. Wang, Y. Chang, Y. Li, Y. Wang, and D. Cao, "4d mmwave radar for autonomous driving perception: A comprehensive survey," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 4, pp. 4606–4620, 2024.

[2] Z. Han, J. Wang, Z. Xu, S. Yang, L. He, S. Xu, J. Wang, and K. Li, "4d millimeter-wave radar in autonomous driving: A survey," *arXiv preprint arXiv:2306.04242*, 2023.

[3] X. Peng, M. Tang, H. Sun, K. Bierzynski, L. Servadei, and R. Wille, "4d mmwave radar for sensing enhancement in adverse environments: Advances and challenges," *arXiv preprint arXiv:2503.24091*, 2025.

[4] 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)*, 2020, pp. 152–159.

[5] J. Zhang, H. Zhuge, Z. Wu, G. Peng, M. Wen, Y. Liu, and D. Wang, "4DRadarSLAM: A 4D Imaging Radar SLAM System for Large-scale Environments based on Pose Graph Optimization," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, May 2023, pp. 8333–8340.

[6] Y. Zhuang, B. Wang, J. Huai, and M. Li, "4D iRIOM: 4D Imaging Radar Inertial Odometry and Mapping," *IEEE Robotics and Automation Letters*, vol. 8, no. 6, pp. 3246–3253, June 2023.

[7] X. Li, H. Zhang, and W. Chen, "4D Radar-Based Pose Graph SLAM With Ego-Velocity Pre-Integration Factor," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 5124–5131, Aug. 2023.

[8] X. Wu, Y. Chen, Z. Li, Z. Hong, and L. Hu, "Efear-4d: Ego-velocity filtering for efficient and accurate 4d radar odometry," *IEEE Robotics and Automation Letters*, vol. 9, no. 11, pp. 9828–9835, 2024.

[9] D. C. Herraez, M. Zeller, D. Wang, J. Behley, M. Heidingsfeld, and C. Stachniss, "Rai-slam: Radar-inertial slam for autonomous vehicles," *IEEE Robotics and Automation Letters*, vol. 10, no. 6, pp. 5257–5264, 2025.

[10] J. Zhang, S. Singh, *et al.*, "Loam: Lidar odometry and mapping in real-time." in *Robotics: Science and systems*, vol. 2, no. 9. Berkeley, CA, 2014, pp. 1–9.

[11] T. Shan and B. Englot, "Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain," in *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2018, pp. 4758–4765.

[12] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, "Fast-lio2: Fast direct lidar-inertial odometry," *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2053–2073, 2022.

[13] L. Wang, X. Zhang, C. Li, D. He, Y. Pan, and J. Yi, "Super-lio: A robust and efficient lidar-inertial odometry system with a compact mapping strategy," *IEEE Robotics and Automation Letters*, vol. 11, no. 3, pp. 2666–2673, 2026.

[14] I. Vizzo, T. Guadagnino, B. Mersch, L. Wiesmann, J. Behley, and C. Stachniss, "Kiss-icp: In defense of point-to-point icp–simple, accurate, and robust registration if done the right way," *IEEE Robotics and Automation Letters*, vol. 8, no. 2, pp. 1029–1036, 2023.

[15] B. Hexsel, H. Vhavle, and Y. Chen, "Dicp: Doppler iterative closest point algorithm," *arXiv preprint arXiv:2201.11944*, 2022.

[16] D. C. Herraez, M. Zeller, L. Chang, I. Vizzo, M. Heidingsfeld, and C. Stachniss, "Radar-only odometry and mapping for autonomous vehicles," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 10 275–10 282.

[17] Q. Huang, Y. Liang, Z. Qiao, S. Shen, and H. Yin, "Less is more: Physical-enhanced radar-inertial odometry," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 15 966–15 972.

[18] D. Adolfsson, M. Karlsson, V. Kubelka, M. Magnusson, and H. Andreasson, "Tbv radar slam – trust but verify loop candidates," *IEEE Robotics and Automation Letters*, vol. 8, no. 6, pp. 3613–3620, 2023.

[19] M. Hilger, V. Kubelka, D. Adolfsson, H. Andreasson, and A. J. Lilienthal, "Towards introspective loop closure in 4d radar slam," *arXiv preprint arXiv:2404.03940*, 2024.

[20] M. Hilger, V. Kubelka, D. Adolfsson, R. Becker, H. Andreasson, and A. J. Lilienthal, "Introspective loop closure for slam with 4d imaging radar," *arXiv preprint arXiv:2503.02383*, 2025.

[21] Z. Liu and F. Zhang, "BALM: Bundle Adjustment for Lidar Mapping," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3184–3191, Apr. 2021.

[22] Z. Liu, X. Liu, and F. Zhang, "Efficient and consistent bundle adjustment on lidar point clouds," *IEEE Transactions on Robotics*, vol. 39, no. 6, pp. 4366–4386, 2023.

[23] X. Liu, Z. Liu, F. Kong, and F. Zhang, "Large-scale lidar consistent mapping using hierarchical lidar bundle adjustment," *IEEE Robotics and Automation Letters*, vol. 8, no. 3, pp. 1523–1530, 2023.

[24] D. Borrmann, J. Elseberg, K. Lingemann, A. Nüchter, and J. Hertzberg, "Globally consistent 3d mapping with scan matching," *Robotics and Autonomous Systems*, vol. 56, no. 2, pp. 130–142, 2008.

[25] D. Adolfsson, M. Magnusson, A. Alhashimi, A. J. Lilienthal, and H. Andreasson, "Lidar-level localization with radar? the cfear approach to accurate, fast, and robust large-scale radar odometry in diverse environments," *IEEE Transactions on Robotics*, vol. 39, no. 2, pp. 1476–1495, 2023.

[26] C. Schöffmann, B. Ubezio, C. Böhm, S. Mühlbacher-Karrer, and H. Zangl, "Virtual radar: Real-time millimeter-wave radar sensor simulation for perception-driven robotics," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4704–4711, 2021.

[27] J. Zhang, H. Zhuge, Y. Liu, G. Peng, Z. Wu, H. Zhang, Q. Lyu, H. Li, C. Zhao, D. Kircali, S. Mharolkar, X. Yang, S. Yi, Y. Wang, and D. Wang, "NTU4DRadLM: 4D Radar-Centric Multi-Modal Dataset for Localization and Mapping," in *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, Sept. 2023, pp. 4291–4296.

[28] M. Choi, S. Yang, S. Han, Y. Lee, M. Lee, K. H. Choi, and K.-S. Kim, "MSC-RAD4R: ROS-Based Automotive Dataset With 4D Radar," *IEEE Robotics and Automation Letters*, vol. 8, no. 11, pp. 7194–7201, Nov. 2023.

[29] H. Kim, M. Jung, C. Noh, S. Jung, H. Song, W. Yang, H. Jang, and A. Kim, "HeRCULES: Heterogeneous Radar Dataset in Complex Urban Environment for Multi-session Radar SLAM," Feb. 2025.

[30] M. Grupp, "evo: Python package for the evaluation of odometry and slam." https://github.com/MichaelGrupp/evo, 2017.