

CoLRIO: LiDAR-Ranging-Inertial Centralized State Estimation for Robotic Swarms

Shipeng Zhong, Hongbo Chen, Yuhua Qi*, Dapeng Feng,
Zhiqiang Chen, Jin Wu, Weisong Wen and Ming Liu

Abstract— Collaborative state estimation using different heterogeneous sensors is a fundamental prerequisite for robotic swarms operating in GPS-denied environments, posing a significant research challenge. In this paper, we introduce a centralized system to facilitate collaborative LiDAR-ranging-inertial state estimation, enabling robotic swarms to operate without the need for anchor deployment. The system efficiently distributes computationally intensive tasks to a central server, thereby reducing the computational burden on individual robots for local odometry calculations. The server back-end establishes a global reference by leveraging shared data and refining joint pose graph optimization through place recognition, global optimization techniques, and removal of outlier data to ensure precise and robust collaborative state estimation. Extensive evaluations of our system, utilizing both publicly available datasets and our custom datasets, demonstrate significant enhancements in the accuracy of collaborative SLAM estimates. Moreover, our system exhibits remarkable proficiency in large-scale missions, seamlessly enabling ten robots to collaborate effectively in performing SLAM tasks. In order to contribute to the research community, we will make our code open-source and accessible at <https://github.com/PengYu-team/Co-LRIO>.

I. INTRODUCTION

Autonomous robotic swarms have gained significant traction in various public applications, such as exploration, inspection, search-and-rescue, healthcare, housework, and logistics [1], [2]. In these applications, accurate estimation the six-degree-of-freedom states of individual robots is pivotal, as it underpins feedback control, obstacle avoidance, and path planning. While state estimation and mapping for single-robot systems can leverage state-of-the-art SLAM techniques, such as Visual Inertial Odometry (VIO) [3], [4] and LiDAR Inertial Odometry (LIO) [5]–[7], the transition to cooperative localization for robotics swarms introduces unique complexities, relying on onboard sensors and intra-swarm communication.

With the growing need for cooperative localization, various algorithms have been developed, some relying on ex-

This research was funded by Research on Path Planning Algorithm of Swarm Unmanned System Based on Deep Reinforcement Learning of China University Industry, Education and Research Innovation Fund No. 2021ZYA11010. Corresponding author is Yuhua Qi.

S. Zhong, H. Chen, Y. Qi, D. Feng and Z. Chen are with the School of Systems Science and Engineering, Sun Yat-sen University, Guangzhou, China (e-mail: zhongshp5, fengdp5, chenzhq56@mail2.sysu.edu.cn, chen-hongbo, qiyh8@mail.sysu.edu.cn).

J. Wu and M. Liu are with Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: jin_wu_uestc@hotmail.com, eelium@ust.hk).

W. Wen is with Department of Aeronautical and Aviation Engineering, Hong Kong Polytechnic University, Hong Kong (e-mail: Wel-sen.wen@polyu.edu.hk).

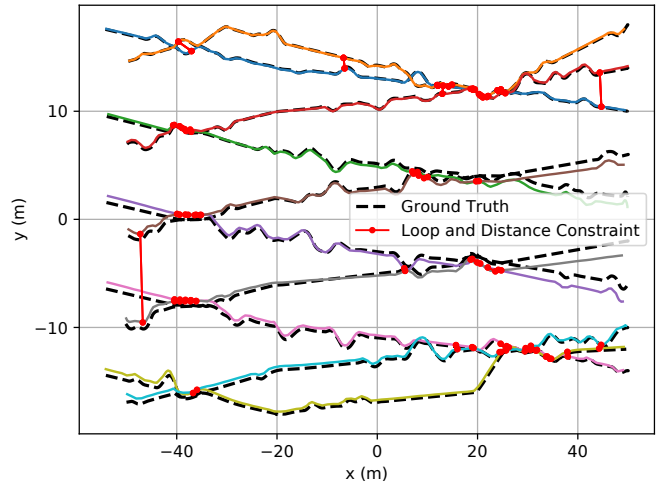


Fig. 1. Simulation results of CoLRIO estimate from 10 drones. The total trajectory length exceeds 1000 m; the black dotted line indicates the ground truth, and the loop and distance constraints are denoted by red lines.

ternal devices like motion capture [8], fixed Ultra-WideBand (UWB) anchors [9], or Global Position System (GPS) [10] to estimate robot states in a common reference frame. However, these solutions are not always available, often constrained by factors such as urban environments limiting GPS reliability or the need for infrastructure deployment (e.g., UWB anchors), restricting their versatility.

An alternative approach to circumvent these limitations involves establishing constraints between robots using vision detection or place recognition [11], [12]. These constraints can arise when sensors on multiple robots simultaneously observe shared landmarks or when one robot observes another within its sensor's field of view. However, shared observations necessitate significant view overlap between neighboring robots, constraining them to remain close and limiting the swarm's flexibility.

This paper introduces a real-time and robust centralized state estimation system tailored for a team of robots equipped with LiDAR, IMU, and UWB sensor suites, enabling collaborative localization without deploying any anchors. Each robot employs direct LIO using the intensity-aided Fast-GICP method for efficient scan-to-map registration. Concurrently, a central server gathers pairwise distance measurements, conducts inter-robot loop closure detection, carries out robust joint optimization, and subsequently furnishes the robots with optimized poses as feedback. Extensive performance evaluations of our system are conducted using

publicly available datasets and our custom datasets, which reveal improvements in the accuracy of collaborative SLAM estimates. Notably, the proposed system exhibits support for collaborative SLAM involving up to 10 robots (Fig. 1), underlining its efficacy in swarm missions.

To summarize, the contributions of this work include: (i): A tightly coupled LiDAR inertial odometry built atop fixed-lag smoothing and intensity-aided direct raw points registration. (ii): An online, robust, and centralized LiDAR-ranging-inertial state estimation system for the robotic swarm enables co-localization without deploying anchors. (iii): Extensive evaluation of the proposed framework in datasets and simulation, including performance, communication and scalability, is conducted. The source code and custom datasets are open to the public.

II. RELATED WORK

A. LiDAR Inertial Odometry

LiDAR odometry is typically required to operate in real-time. One of the renowned feature-based algorithms in this domain is LOAM [13], which extracts edge and plane features based on local smoothness. Several works maintain a framework similar to LOAM, such as LIO-SAM [5], LRIO [7], and LiLi-OM [14]. These approaches incorporate IMU to compensate for motion distortion and provide an initial estimate for scan matching. However, achieving real-time performance often comes at the cost of maintaining only a small local map, which can lead to drift over time. Consequently, a map refining process, such as factor graph optimization, becomes necessary. FAST-LIO2 [6] introduces an ESIKF and incremental kd-tree to enhance computational efficiency, enabling direct and real-time scan-to-map registration. Compared to the aforementioned feature-based methods, FAST-LIO2 offers improved odometry accuracy while eliminating the need for parameter tuning for feature extraction. Nevertheless, maintaining the incremental map can be challenging in multi-robot scenarios due to the introduction of numerous inter-robot measurements. LION [15] also directly registers raw points to the map, which is very similar to the idea of GICP [16] and Normal Distribution Transformation (NDT) [17]. However, the inefficiencies of GICP [16] and the limited stability of NDT are issues that demand attention. To address these challenges, Fast-GICP [18] adopts the multithread technology to alleviate the computational load. Our work inherits this approach, aiming to improve efficiency and stability in front-end odometry.

B. Multi-robot Localization and Mapping

Estimating the relative pose between robots is a critical task, primarily aimed at compensating for odometry drift without the reliance on external positioning infrastructure. Using vision detection [19] has shown the capability to achieve centimeter-level accuracy in determining relative poses. Nevertheless, it suffers from limitations related to onboard camera visibility and restricted field-of-view (FoV), which can constrain the trajectories and formation between robots. An efficient and widely used approach involves place

recognition through global or object-based features. Visual-based collaborative Simultaneous Localization and Mapping (CSLAM) systems [20]–[22] share the binary descriptor (e.g., ORB) of the keyframes and employ a bag-of-words (BoWs) approach to find multi-robot loop closures. Similarly, laser-based CSLAM systems [23]–[25] extract global features for loop closure detection. LAMP [26] use TEASER [27] and GICP algorithm [16] for relative transform estimation. These approaches rely on having sufficient overlapping features within the robots' visual range, which can restrict swarm formation and orientation, rendering them most effective in feature-rich environments. A different approach, presented in [7], [28], [29], leverages UWB measurements with a fixed anchor to eliminate odometry drift. While this strategy demonstrates promise, it necessitates additional infrastructure deployment, limiting its practicality in real-world scenarios. Moreover, other works [12], [30]–[32] fuse odometry, particularly VIO, with UWB distance measurements to estimate relative robot states. Notably, in the case of Omni-swarm [12], vision-based detection and place recognition are also incorporated for multi-robot state estimation. Unlike place recognition methods, the approach in [30]–[32] does not mandate trajectory overlaps, granting it greater mission flexibility. However, achieving global consistency and robust optimization under these methods remains challenging, mainly when dealing with line-of-sight limitations inherent to distance measurements. This work introduces lightweight global descriptors and distance measurements for efficient loop closure detection, and robust optimization ensures global consistency, making it a promising multi-robot state estimation solution in challenging environments.

III. RELATIVE STATE ESTIMATION FRAMEWORK

The proposed system architecture comprises two primary components: a central server and a team of robots, as visually represented in Fig. 2. The robot with a unique identification in the team is designed to join or exit the collaborative optimization freely. Collecting raw data from a LiDAR, an IMU, and a UWB radio, each robot performs a LIO algorithm to estimate its state. In our implementation, we employ a tailored adaptation of Fast-GICP method for the scan-to-map matching process, finding the point's correspondence with the help of its intensity. The odometry utilizes a fixed-lag smoother, a popular method for the visual SLAM, to solve the state estimation problem efficiently in a sliding window.

Processing every LiDAR frame for computing and incorporating factors into the back-end factor graph proves to be computationally and transmission demanding for the server. To mitigate this, we adopt the concept of keyframe selection for real-time nonlinear optimization, striking a balance between map density and memory consumption by maintaining a relatively sparse factor graph. Keyframes are succinctly described using the lightweight global descriptor Scan-Context++ [33]. These keyframes, relative odometry, and distance measurements are transmitted to the server. The Multi-map Loop Closure module detects overlaps in the trajectories of robots and reports loop closure factors. The

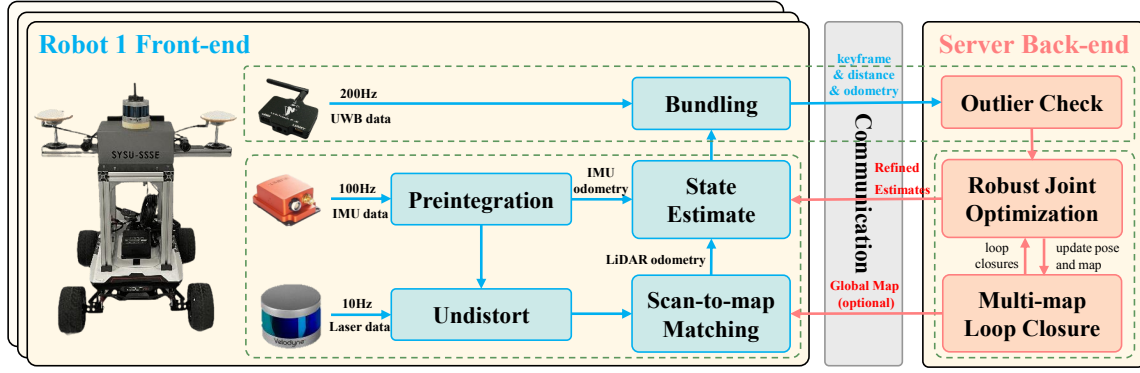


Fig. 2. The system structure of CoLRIO. In the front end, the preintegration and direct scan-to-map matching odometry are further jointly optimized with fixed-lag smoothness. Keyframes generated by LIO, which include UWB measurements, are transmitted to the server. The server collects all measurements, executing robust joint optimization to achieve swarm state estimations. Finally, the estimates are returned to the front end for further processing.

distance measurements are further validated against odometry poses to identify and reject outliers. Odometry, distance, and loop closure factors perform robust joint optimization utilizing Graduated Non-Convexity (GNC) [34] to obtain the trajectory estimates. The process for generating the above factors is described in the following section.

A. Single-robot Front-end

1) *IMU Preintegration Factor*: The IMU preintegration factor deals with raw IMU measurements, encompassing angular velocity $\tilde{\mathbf{w}}$ and linear acceleration $\tilde{\mathbf{a}}$ at time t . The equations governing these measurements are as follows:

$$\begin{aligned} \tilde{\mathbf{w}}_t &= \mathbf{w}_t + \mathbf{b}_t^{\mathbf{w}} + \mathbf{n}^{\mathbf{w}}, & \mathbf{b}_t^{\mathbf{w}} &= \mathbf{n}^{\mathbf{b}^{\mathbf{w}}} \\ \tilde{\mathbf{a}}_t &= \mathbf{R}_t^T (\mathbf{a}_t - \mathbf{g}) + \mathbf{b}_t^{\mathbf{a}} + \mathbf{n}^{\mathbf{a}}, & \mathbf{b}_t^{\mathbf{a}} &= \mathbf{n}^{\mathbf{b}^{\mathbf{a}}} \end{aligned} \quad (1)$$

where \mathbf{w} and \mathbf{a} denote the true angular velocity and linear acceleration, $\mathbf{n}^{\mathbf{w}}$ and $\mathbf{n}^{\mathbf{a}}$ are zero-mean Gaussian white noise. The gyroscope bias $\mathbf{b}^{\mathbf{w}}$ and accelerometer bias $\mathbf{b}^{\mathbf{a}}$ are modeled as random walks, driving by the white Gaussian noises $\mathbf{n}^{\mathbf{b}^{\mathbf{w}}}$ and $\mathbf{n}^{\mathbf{b}^{\mathbf{a}}}$, respectively. We compute the relative body motion between two timestamps, similar to the IMU preintegration method introduced in [35]. With a time offset Δt between two frames, the IMU preintegration factor \mathbf{r}^I can be expressed as:

$$\mathbf{r}_{t_i, t_j}^I = \begin{pmatrix} \Delta \hat{\mathbf{R}}_{t_i, t_j}^T (\mathbf{R}_{t_i}^T \mathbf{R}_{t_j}) \\ \mathbf{R}_{t_i}^T (\mathbf{v}_{t_j} - \mathbf{v}_{t_i} - \mathbf{g} \Delta t) - \Delta \hat{\mathbf{v}}_{t_i, t_j} \\ \mathbf{R}_{t_i}^T (\mathbf{p}_{t_j} - \mathbf{p}_{t_i} - \mathbf{v}_{t_i} \Delta t - \frac{1}{2} \mathbf{g} \Delta t^2) - \Delta \hat{\mathbf{p}}_{t_i, t_j} \end{pmatrix}, \quad (2)$$

where $\hat{\mathbf{R}}$, $\hat{\mathbf{p}}$, and $\hat{\mathbf{v}}$ indicates the estimated rotation, translation, and velocity.

2) *LiDAR Odometry Factor*: Suppose a new LiDAR scan \mathbf{S} arrives at time t . The previously preintegrated IMU data prior to time t can be employed to mitigate distortions in the scan caused by motion. It is worth noting that this new scan \mathbf{S} is represented in the body frame. The process of generating a LiDAR odometry factor is described as follows:

Voxel map maintaining. Rather than optimizing the transformation between consecutive scans, we transform historical scans into the map frame using corresponding poses and compute the scan-to-map transformation for each scan. Given the high frequency of map reconstruction caused by

global estimate refining, we employ a classic kd-tree-based map instead of an incremental map. Additionally, to save computational resources, keyframes are added to strike a balance between map density and memory consumption. In our work, the criteria for adding a new keyframe are set at position and rotation change thresholds of 1 m and 0.2 rad, respectively. We downsample the map at a resolution of 0.4 meters to manage computational resources effectively.

Scan-to-map matching. We match a newly arrived scan \mathbf{S} to the map \mathbf{M} via raw point registration using the Fast-GICP method, chosen for its computational efficiency and robustness across various challenging environments. For rapid convergence, we use the predicted robot motion from IMU preintegration as the initial transformation for scan-to-map matching. For each point \mathbf{n} in scan \mathbf{S} , we identify its corresponding point \mathbf{m} in map \mathbf{M} based on both position and intensity. Specifically, inspired by the two-stage point neighbor search approach proposed in [36], we first perform a position-based k-nearest-neighbor search. Subsequently, an intensity-based k-nearest-neighbor search is conducted using the neighbor points obtained from the previous step. This dual-stage process allows us to further gather points with similar geometry and materials. Furthermore, we compute the covariance matrix Σ of the corresponding point using the neighbor points. The transformation error between the scan \mathbf{S} and the map \mathbf{M} can be computed as follows:

$$e = \frac{1}{N} \sum_{i=1}^N w_i \|\mathbf{R} \cdot \mathbf{n}_i + \mathbf{p} - \mathbf{m}_i\|^2, \quad (3)$$

where N is the number of the corresponding points set, \mathbf{R} and \mathbf{p} are the rotation matrix and translation vector that transform the source point cloud to align with the target point cloud, \mathbf{m}_i and \mathbf{n}_i are the i -th of corresponding points, respectively, and $w_i = (\Sigma_{\mathbf{m}_i} + \mathbf{T} \Sigma_{\mathbf{n}_i} \mathbf{T}^T)^{-1}$ is the weight based on the transformation $\mathbf{T} = [\mathbf{R} | \mathbf{p}]$ and covariance matrix Σ of the points. We compute the final transformation \mathbf{T}_f , minimizing the transformation error e . Subsequently, we obtain the relative transformation $\Delta \mathbf{X}_{t_i, t_j} = \mathbf{X}_{t_i}^{-1} \mathbf{X}_{t_j} = \mathbf{X}_{t_i}^{-1} \mathbf{T}_f$ between robot state \mathbf{X}_{t_i} and \mathbf{X}_{t_j} , with odometry measurements modeled with Gaussian noise, denoted as $\mathbf{Z}_{t_i, t_j}^O = \Delta \mathbf{X}_{t_i, t_j} + \eta(0, \Sigma_O)$.

Resulting in the following odometry factor \mathbf{r}^O :

$$\mathbf{r}_{i,t_j}^O = (\mathbf{Z}_{i,t_j}^O)^{-1}(\hat{\mathbf{X}}_i)^{-1}\hat{\mathbf{X}}_{t_j}, \quad (4)$$

where $\hat{(\cdot)}$ indicates the estimated state.

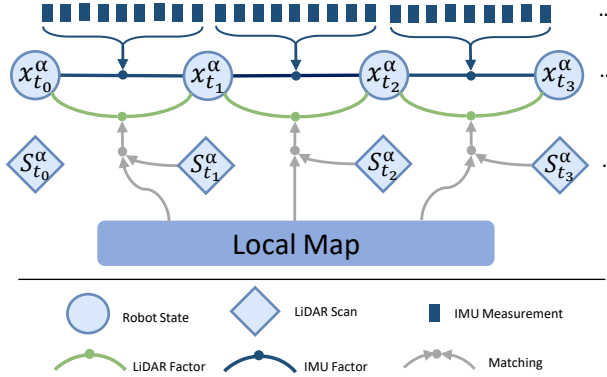


Fig. 3. The structure of front-end odometry, fusing LiDAR and IMU data.

As depicted in Fig. 3, the front-end fuses different measurements in a factor graph optimization framework, including LiDAR odometry factors, IMU factors, and a prior factor (including the initial pose and the refined estimates). We formulate a Nonlinear Least-Squares (NLS) problem:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left\{ \sum_{\mathbf{O}} \|\mathbf{r}^O\|^2 + \sum_{\mathbf{I}} \|\mathbf{r}^I\|^2 + \|\mathbf{r}^P\|^2 \right\}, \quad (5)$$

where $\mathbf{O/I}$ is the set of all odometry/IMU measurements.

Since adding every odometry factor to the graph is computationally infeasible, old factors in the factor graph are marginalized to bound the problem size, which is solved through the repeated linearization of factors. Utilizing probabilistic marginalization ensures that information on the active states is well-preserved in the optimization, even as measurements and old states are removed from the sliding window. In the experiments, we achieve multi-source data joint optimization using an iSAM2-based fixed-lag smoother with a sliding window containing ten keyframes.

B. Swarm Localization

1) *Loop Closure Detection*: To establish additional constraints within and across the robots' trajectories, we employ loop-closure detection, utilizing the global descriptor ScanContext++ [33]. The server maintains a global database of the keyframe descriptors transmitting from robots to enable inter-robot loop detection. New loop closures are detected by querying the database for nearest neighbor candidates, with the candidates subjected to scan-to-map matching using FastGICP. Similar to equation (3), the relative pose of robot α and β , indicated by ${}_\alpha\mathbf{X}_{t_i}$ and ${}_\beta\mathbf{X}_{t_m}$, in case of loop detection between different robots or within the same trajectory, respectively, is optimized by minimizing the transformation error. The loop closure is accepted if the fitness score does not exceed the threshold. Subsequently, we obtain the relative transformation ${}_{\alpha,\beta}\mathbf{Z}_{i,t_m}$ between ${}_\alpha\mathbf{X}_{t_i}$ and ${}_\beta\mathbf{X}_{t_m}$, and the loop closure measurements modeled with a Gaussian noise,

denoted as ${}_{\alpha,\beta}\mathbf{Z}_{i,t_m}^L = ({}_\alpha\mathbf{X}_{t_i})^{-1}{}_\beta\mathbf{X}_{t_m} + \eta(0, \Sigma_L)$. Resulting in the following loop closure factor \mathbf{r}^L :

$${}_{\alpha,\beta}\mathbf{r}_{i,t_m}^L = ({}_{\alpha,\beta}\mathbf{Z}_{i,t_m}^L)^{-1}({}_\alpha\hat{\mathbf{X}}_{t_i})^{-1}{}_\beta\hat{\mathbf{X}}_{t_m}. \quad (6)$$

2) *Pairwise Distance Measurement*: From the UWB modules, the relative distances between each pair of the UWB nodes can be obtained, represented as the UWB distance measurement \mathbf{Z}_t^U at time t , modeled with Gaussian noise, denoted as ${}_{\alpha,\beta}\mathbf{Z}_t^U = \|\alpha\mathbf{p}_t - \beta\mathbf{p}_t\|^2 + \eta(0, \Sigma_U)$. The occlusion of obstacles often results in significant outliers in the UWB measurements. To assess the consistency of the distance measurements with the IMU odometry produced by the front end, we consider any measurement that exceeds the distance outlier threshold, denoted as ${}_{\alpha,\beta}\mathbf{Z}_t^U - ({}_\alpha\hat{\mathbf{p}}_t - {}_\beta\hat{\mathbf{p}}_t) < \sigma$ as an outlier. Note that the position ${}_\alpha\hat{\mathbf{p}}_t$ and ${}_\beta\hat{\mathbf{p}}_t$ of IMU odometry are in the same frame. We set the distance outlier threshold to 0.2 m, which is twice the expected error of the UWB measurements. The relative distance factor \mathbf{r}^{uwb} as follows:

$${}_{\alpha,\beta}\mathbf{r}_t^U = {}_{\alpha,\beta}\mathbf{Z}_t^U - \|\alpha\hat{\mathbf{p}}_t - \beta\hat{\mathbf{p}}_t\|^2. \quad (7)$$

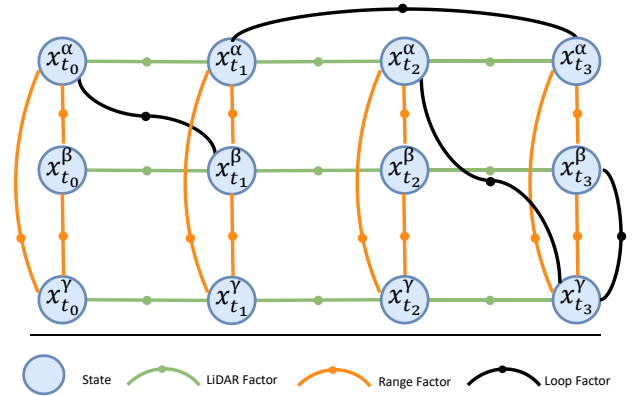


Fig. 4. The structure of the CoLRIO back-end. LiDAR odometry, range, and loop closure factors are introduced to global optimization.

3) *Global Optimization*: As illustrated in Fig. 4, the factors that need to be solved consists of the above LiDAR odometry factors \mathbf{r}^O , relative distance factors \mathbf{r}^U , loop closure factors \mathbf{r}^L and a prior factor \mathbf{r}^P (the initial pose of one of the robots). We formulate the following NLS problem:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left\{ \sum_{\mathbf{O}} \|\mathbf{r}^O\|^2 + \sum_{\mathcal{U}} \|\mathbf{r}^U\|^2 + \sum_{\mathcal{L}} \|\mathbf{r}^L\|^2 + \|\mathbf{r}^P\|^2 \right\}, \quad (8)$$

where $\mathcal{O/U/L}$ is the set of all key odometry/distance/loop closure measurements. To enhance the robustness and reliability of our server back-end, we have seamlessly integrated two outlier rejection mechanisms: Pairwise Consistent Measurement set maximization (PCM) [37] and GNC [34]. PCM scrutinizes inter-robot loop closures for consistency with other odometry. Meanwhile, GNC works alongside Levenberg-Marquardt to execute an outlier-robust factor graph optimization, providing trajectory estimates and making inlier/outlier determinations on the inter-robot (loops and distances) measurements not rejected by PCM. Both PCM and GNC are implemented with the GTSAM library.

IV. EXPERIMENTS

1) *Setup*: All experimental results presented in this paper were derived from performance comparisons conducted on five indoor and outdoor sequences of the S3E dataset [38] and six datasets collected on SYSU’s eastern campus. We leverage Robot Operating System 2 (ROS2) [39] and a PC equipped with an Intel i7-10700K CPU and 32GB DDR4 RAM for these experiments. Table I summarizes the critical characteristics of the datasets utilized in our subsequent experiments. Inheriting the hardware of three ground vehicles in the S3E dataset, we added a LinkTrack P-B for inter-robot distance measurement, upgraded the original 1Hz Real-time Kinematic (RTK) system to a high-frequency 100Hz CHCNV CGI-610 GNSS unit as the ground truth to ensure the accuracy and reliability of the ground truth data (especially elevation). Note that AMOVLAB provides some devices for ground vehicles. Moreover, the CGI-610 served as an external trigger, enabling precise synchronization of the Xsens MTi-30 IMU and Velodyne VLP-16 Puck LiDAR systems. Additionally, it was instrumental in maintaining synchronized system time among the robotic platforms via GPS timing and synchronization. To evaluate the trajectories, we employed ground truth to compute the Absolute Translation Error (ATE) using evo toolbox. For indoor S3E sequences, only the initial and final positions are compared.

TABLE I
DETAILS OF S3E AND OUR CAMPUS DATASETS

Datasets	Sensor	Trajectory Length [m]			Size [GB]	
		Robot 1	Robot 2	Robot 3		
S3E	<i>Square_1</i>	LVI	455.6	454.4	458.2	19.1
	<i>College</i>	LVI	920.5	995.9	1072.3	31.6
	<i>Playground_2</i>	LVI	277.7	319.9	474.0	6.7
	<i>Laboratory_1</i>	LVI	146.7	161.6	141.2	9.4
	<i>Laboratory_2</i>	LVI	213.5	197.9	160.2	6.7
Our dataset	<i>Archway</i>	LRI	487.4	569.6	563.8	8.5
	<i>Rotation</i>	LRI	84.7	91.3	110.7	1.8
	<i>Small_Loop</i>	LRI	521.6	523.9	520.3	9.3
	<i>Large_Loop</i>	LRI	1938.6	1934.2	1950.1	30.9
	<i>Tunnel</i>	LRI	521.9	502.4	501.1	8.2
	<i>College</i>	LRI	983.3	967.6	986.9	17.5

† LVI indicates LiDAR-Visual-Inertial and LRI indicates LiDAR-Ranging-Inertial.

2) *Dataset Experiments*: To comprehensively evaluate the accuracy and robustness of our front-end odometry, we conducted a comparative analysis, pitting the proposed CoLRIO front end against state-of-the-art odometry such as LIO-SAM and FAST-LIO2 across diverse datasets. Note that all front-ends utilize the centralized server to detect only intra-robot loop closure to reduce odometry drift and optimize the trajectory with GNC. The evaluation results, specifically the ATE, are presented in Table II. Fig. 5(a) and 5(b) illustrate the trajectories achieved by different front ends. Notably, the *Archway*, *Tunnel*, and *S3E_Square* sequences do not include intra-robot loop closures. From Table II, it’s evident that the CoLRIO front end consistently outperforms single-robot odometry methods, including LIO-SAM and FAST-LIO2, thereby demonstrating its effective-

ness and stability. Thanks to its efficient direct registration utilizing raw point clouds, the front-end odometry can seamlessly adapt to challenging environmental conditions. This enables enhanced precision in state estimation even during rapid motion (*S3E_Playground_2*), extensive angular rotation (*Rotation*), long-distance traversal (*Large_Loop*), and indoor (*S3E_Laboratory_1* and *S3E_Laboratory_2*) settings. Additionally, regarding computational cost, with 4 parallel threads for the GICP algorithm, each robot performs our front-end with a maximum CPU usage of 17.4%, taking an average of 50.03ms for each frame in the *Large_Loop* sequence, highlighting its effectiveness with real-time performance. In summary, our front end exhibits efficient and precise localization performance, laying the foundation for joint optimization in the robotic swarms scenarios.

In order to further assess the precision and resilience of our entire SLAM system, we performed a comparative analysis by benchmarking the proposed CoLRIO method against Swarm-SLAM and DiScO-SLAM across various datasets. Fig. 5(c) and 5(d) depict the trajectories achieved by different C-SLAM systems. Throughout our experiments, we observed that DiScO-SLAM and Swarm-SLAM exhibited satisfactory performance in specific sequences. However, they encountered difficulties in merging all sub-maps across most sequences. Based on experimental observations, the limitations of DiScO-SLAM can be attributed to deficiencies in their loop closure searching module and the lack of robustness in the front-end LIO-SAM. In contrast, Swarm-SLAM exhibits elevated CPU usage in front-end RTAB-Map and loop closure processes. Insufficient computational resources may result in non-real-time odometry localization (less than 1Hz) and significantly impact the performance of the loop closure module. Unlike DiScO-SLAM and Swarm-SLAM, our CoLRIO approach demonstrates very competitive performance on most sequences, showcasing its robustness for different environments. Comparing CoLRIO to the front-end, we observed performance improvements in the *Large_Loop*, *College* (characterized by long duration), and *Tunnel* (marked by severe perceptual aliasing). These advancements were primarily attributed to leveraging distance measurements from UWB radio, facilitating robust optimization, and effectively eliminating outlier loop closure measurements. In the other three sequences, all robots followed forward trajectories, eventually meeting at specific points (*Archway*) or returning to their initial positions after a short journey (*Small_Loop* and *Rotation*). Although our proposed system merged these trajectories, it could not mitigate front-end drift further due to insufficient inter-robot measurements. In short, the proposed CoLRIO framework demonstrates its effectiveness in merging sub-maps among robotic teams and robustness in challenging environments with the help of distance data.

3) *Communication*: To quantify the bandwidth requirements of the proposed system, we conducted a comprehensive network traffic analysis on both the S3E and our campus datasets (using the Velodyne VLP-16). The average bandwidth utilization is presented in Table III. In this configuration, each robot communicates with the central server

TABLE II

ABSOLUTE TRANSLATION ERROR (METER) OF THE S3E SEQUENCES AND OUR CAMPUS DATASETS

Configure	Datasets	Our campus dataset					S3E dataset					
		<i>Archway</i>	<i>Rotation</i>	<i>Small_Loop</i>	<i>Large_Loop</i>	<i>Tunnel</i>	<i>College</i>	<i>Playground_2</i>	<i>Square_1</i>	<i>College</i>	<i>Laboratory_1</i>	<i>Laboratory_2</i>
LIO-SAM		54.22	5.71	1.01	131.52	3.85	49.24	24.81	0.94	3.05	3.99	11.06
FAST-LIO2		4.55	0.35	1.83	35.00	4.35	5.58	20.72	3.00	8.30	8.30	9.13
our front-end		1.63	0.26	0.92	6.16	3.53	0.81	0.48	0.73	1.01	0.46	0.57
Swarm-SLAM		6.15	3.72	3.20	22.42	10.54	7.93	1.45	4.20	3.57	4.76	8.97
DiSCo-SLAM		7.43	5.79	1.03	23.38	7.71	12.02	3.82	1.03	1.76	1.20	2.33
CoLRIO w/o UWB		1.64	0.26	0.94	1.97	5.07	1.08	0.63	0.75	0.98	0.46	0.18
our CoLRIO		1.59	0.26	0.88	1.47	3.54	0.77	0.63	0.75	0.98	0.46	0.18

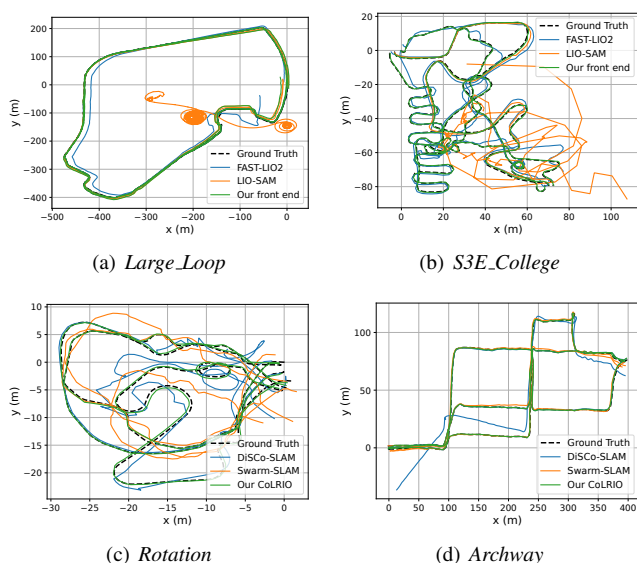


Fig. 5. Different methods' trajectory estimates on various dataset sequences compared with ground truth.

TABLE III

NETWORK TRAFFIC OF EACH ROBOT AVERAGED OVER S3E AND CAMPUS SEQUENCES

Datasets	Robot \rightarrow Server \pm Std.	Server \rightarrow Robot \pm Std.
<i>Archway</i>	610.31 \pm 33.35 kBps	0.67 \pm 0.90 kBps
<i>Small_Lopp</i>	663.54 \pm 22.56 kBps	9.89 \pm 7.23 kBps
<i>S3E_College</i>	608.40 \pm 34.29 kBps	11.91 \pm 10.99 kBps
Avg. (3 Seq.)	626.69 kBps	8.63 kBps

at a frequency ranging from 1 to 2 Hz, providing new keyframe data that includes combined odometry and distance measurements. On average, these keyframe messages possess a size of approximately 242.05 kB. Reciprocally, the server communicates with individual robots at a similar frequency range of 1-2 Hz. As delineated in Table III, the network traffic generated by the three robots and conveyed to the server amounts to approximately 600 kBps. The LinkTrack P-B UWB module and a standard WiFi module efficiently manage this data transfer. In contrast, the server's data traffic to each robot is notably lower in our implementation. This is primarily because only necessary global estimates are shared to facilitate the drift correction. It is important to note that transmitted data will increase rapidly for more complex sequences, where the LIO generates a greater number of keyframes. Consequently, the system's bandwidth imposes limitations on the scalability of the robotic team, making

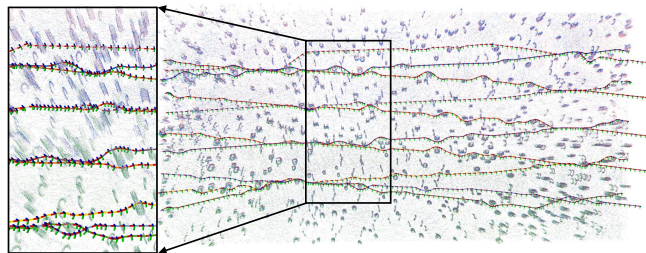


Fig. 6. The reconstructed map of CoLRIO from 10 drones flying across an area full of circle and cylinder obstacles.

it challenging to deploy the system in larger-scale swarms comprising hundreds of robots.

4) *Simulation*: In this experiment, we assess the applicability of our proposed CoLRIO framework in a scenario involving a substantial team of UAVs. For simulation purposes, we employ MARSIM [40], allowing us to orchestrate the flight of 10 drones within an area filled with ring and cylinder obstacles. The trajectories followed by these drones, as depicted in Fig. 1, collectively span over 100 meters and encompass the entire area. The collaborative mapping results produced by CoLRIO are showcased in Fig. 6. The average ATE achieved by CoLRIO, when incorporating distance measurements, stands at a commendable 0.23 m. In contrast, CoLRIO's ATE, without distance measurements, registers at 0.32 m. The experimental results show that the proposed CoLRIO framework can be extended to a certain scale robotic swarm.

V. CONCLUSIONS

This paper introduces CoLRIO, a robust and precise collaborative SLAM framework. CoLRIO empowers multiple robots to collaboratively generate global SLAM estimates from their concurrently collected data in real-time. Our extensive experiments underscore the accuracy and robustness of collaborative SLAM estimates, particularly in large-scale multi-agent missions, where we demonstrate its effectiveness with up to 10 robots simultaneously contributing to the system. In our future work, we aim to extend the scalability of this system, enabling it to gracefully accommodate a multitude of agents, potentially in the hundreds, through innovative strategies such as leveraging MARSIM simulation and employing a distributed or decentralized structure. Additionally, we are committed to enhancing the system's resilience to significant outlier distance and loop measurements using robust incremental optimization.

REFERENCES

- [1] Y. Rizk, M. Awad, and E. W. Tunstel, "Cooperative heterogeneous multi-robot systems: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 2, pp. 1–31, 2019. **1**
- [2] P.-Y. Lajoie, B. Ramtoula, F. Wu, and G. Beltrame, "Towards collaborative simultaneous localization and mapping: a survey of the current research landscape," *Field Robotics*, vol. 2, no. 1, pp. 971–1000, 2022. **1**
- [3] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018. **1**
- [4] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021. **1**
- [5] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 5135–5142. **1, 2**
- [6] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, "Fast-lid2: Fast direct lidar-inertial odometry," *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2053–2073, 2022. **1, 2**
- [7] T.-M. Nguyen, M. Cao, S. Yuan, Y. Lyu, T. H. Nguyen, and L. Xie, "Liro: Tightly coupled lidar-inertia-ranging odometry," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 14 484–14 490. **1, 2**
- [8] A. Jaimes, S. Kota, and J. Gomez, "An approach to surveillance an area using swarm of fixed wing and quad-rotor unmanned aerial vehicles uav(s)," in *2008 IEEE International Conference on System of Systems Engineering*, 2008, pp. 1–6. **1**
- [9] A. Ledergerber, M. Hamer, and R. D'Andrea, "A robot self-localization system using one-way ultra-wideband communication," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 3131–3137. **1**
- [10] S. Moon, Y. Choi, D. Kim, M. Seung, and H. Gong, "Outdoor swarm flight system based on rtk-gps," *Journal of KIISE*, vol. 43, no. 12, pp. 1315–1324, 2016. **1**
- [11] J. P. Queralta, Q. Li, F. Schiano, and T. Westerlund, "Vio-uw-based collaborative localization and dense scene reconstruction within heterogeneous multi-robot systems," in *2022 International Conference on Advanced Robotics and Mechatronics (ICARM)*, 2022, pp. 87–94. **1**
- [12] H. Xu, Y. Zhang, B. Zhou, L. Wang, X. Yao, G. Meng, and S. Shen, "Omni-swarm: A decentralized omnidirectional visual-inertial-uw state estimation system for aerial swarms," *IEEE Transactions on Robotics*, vol. 38, no. 6, pp. 3374–3394, 2022. **1, 2**
- [13] J. Zhang and S. Singh, "LOAM: lidar odometry and mapping in real-time," in *Robotics: Science and Systems X, University of California, Berkeley, USA, July 12-16, 2014*, D. Fox, L. E. Kavraki, and H. Kurniawati, Eds., 2014. **2**
- [14] K. Li, M. Li, and U. D. Hanebeck, "Towards high-performance solid-state-lidar-inertial odometry and mapping," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5167–5174, 2021. **2**
- [15] A. Tagliabue, J. Tordesillas, X. Cai, A. Santamaria-Navarro, J. P. How, L. Carlone, and A.-a. Agha-mohammadi, "Lion: Lidar-inertial observability-aware navigator for vision-denied environments," in *Experimental Robotics*, B. Siciliano, C. Laschi, and O. Khatib, Eds. Cham: Springer International Publishing, 2021, pp. 380–390. **2**
- [16] A. Segal, D. Haehnel, and S. Thrun, "Generalized-icp," in *Robotics: science and systems*, vol. 2, no. 4. Seattle, WA, 2009, p. 435. **2**
- [17] M. Magnusson, "The three-dimensional normal-distributions transform: an efficient representation for registration, surface analysis, and loop detection," Ph.D. dissertation, Örebro universitet, 2009. **2**
- [18] K. Koide, M. Yokozuka, S. Oishi, and A. Banno, "Voxelized gicp for fast and accurate 3d point cloud registration," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 11 054–11 059. **2**
- [19] T. Nguyen, K. Mohta, C. J. Taylor, and V. Kumar, "Vision-based multi-mav localization with anonymous relative measurements using coupled probabilistic data association filter," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 3349–3355. **2**
- [20] P. Schmuck, T. Ziegler, M. Karrer, J. Perraudin, and M. Chli, "Covins: Visual-inertial slam for centralized collaboration," pp. 171–176, 2021. **2**
- [21] M. Ouyang, X. Shi, Y. Wang, Y. Tian, Y. Shen, D. Wang, P. Wang, and Z. Cao, "A collaborative visual slam framework for service robots," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 8679–8685. **2**
- [22] Y. Tian, Y. Chang, F. Herrera Arias, C. Nieto-Granda, J. P. How, and L. Carlone, "Kimera-multi: Robust, distributed, dense metric-semantic slam for multi-robot systems," *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2022–2038, 2022. **2**
- [23] R. Dube, A. Cramariuc, D. Dugas, H. Sommer, M. Dymczyk, J. Nieto, R. Siegwart, and C. Cadena, "Segmap: Segment-based mapping and localization using data-driven descriptors," *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 339–355, 2020. **2**
- [24] Y. Huang, T. Shan, F. Chen, and B. Englot, "Disco-slam: Distributed scan context-enabled multi-robot lidar slam with two-stage global-local graph optimization," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1150–1157, 2022. **2**
- [25] P.-Y. Lajoie and G. Beltrame, "Swarm-slam: Sparse decentralized collaborative simultaneous localization and mapping framework for multi-robot systems," *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 475–482, 2024. **2**
- [26] Y. Chang, K. Ebadi, C. E. Denniston, M. F. Ginting, A. Rosinol, A. Reinke, M. Palieri, J. Shi, A. Chatterjee, B. Morrell *et al.*, "Lamp 2.0: A robust multi-robot slam system for operation in challenging large-scale underground environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9175–9182, 2022. **2**
- [27] H. Yang, J. Shi, and L. Carlone, "Teaser: Fast and certifiable point cloud registration," *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 314–333, 2020. **2**
- [28] H. Zhou, Z. Yao, Z. Zhang, P. Liu, and M. Lu, "An online multi-robot slam system based on lidar/uwb fusion," *IEEE Sensors Journal*, vol. 22, no. 3, pp. 2530–2542, 2022. **2**
- [29] T.-M. Nguyen, S. Yuan, M. Cao, T. H. Nguyen, and L. Xie, "Viral slam: Tightly coupled camera-imu-uwb-lidar slam," *arXiv preprint arXiv:2105.03296*, 2021. **2**
- [30] T. Ziegler, M. Karrer, P. Schmuck, and M. Chli, "Distributed formation estimation via pairwise distance measurements," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3017–3024, 2021. **2**
- [31] J. P. Queralta, Q. Li, F. Schiano, and T. Westerlund, "Vio-uw-based collaborative localization and dense scene reconstruction within heterogeneous multi-robot systems," in *2022 International Conference on Advanced Robotics and Mechatronics (ICARM)*, 2022, pp. 87–94. **2**
- [32] T. H. Nguyen, T.-M. Nguyen, and L. Xie, "Range-focused fusion of camera-imu-uwb for accurate and drift-reduced localization," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1678–1685, 2021. **2**
- [33] G. Kim, S. Choi, and A. Kim, "Scan context++: Structural place recognition robust to rotation and lateral variations in urban environments," *IEEE Transactions on Robotics*, vol. 38, no. 3, pp. 1856–1874, 2022. **2, 4**
- [34] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone, "Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1127–1134, 2020. **3, 4**
- [35] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual-inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2017. **3**
- [36] M. Jung, S. Jung, H. Jang, and A. Kim, "Intensity and ambient enhanced lidar-inertial slam for unstructured construction environment," *The Journal of Korea Robotics Society*, vol. 16, no. 3, pp. 179–188, 2021. **3**
- [37] J. G. Mangelson, D. Dominic, R. M. Eustice, and R. Vasudevan, "Pairwise consistent measurement set maximization for robust multi-robot map merging," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 2916–2923. **4**
- [38] D. Feng, Y. Qi, S. Zhong, Z. Chen, Y. Jiao, Q. Chen, T. Jiang, and H. Chen, "S3e: A large-scale multimodal dataset for collaborative slam," *arXiv preprint arXiv:2210.13723*, 2022. **5**
- [39] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, "Robot operating system 2: Design, architecture, and uses in the wild," *Science Robotics*, vol. 7, no. 66, p. eabm6074, 2022. **5**
- [40] F. Kong, X. Liu, B. Tang, J. Lin, Y. Ren, Y. Cai, F. Zhu, N. Chen, and F. Zhang, "Marsim: A light-weight point-realistic simulator for lidar-based uavs," *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 2954–2961, 2023. **6**