

A Novel Dynamic Localization Graph for Efficient Relative Localization

Gaoming Chen, Kun Song, Wenhong Liu, Wenya Ma, and Zhenhua Xiong,*Member, IEEE*

Abstract—Accurate relative poses are crucial for consistent decision-making in multi-robot systems, which can be realized by Relative Localization (RL). Existing research mainly focuses on determining relative poses through RL without thoroughly assessing the uncertainty in this process. Although some methods have been proposed to increase the RL accuracy, like trajectory repetition, stopping criterion in these efforts are usually empirical since the ground truth of relative pose is unknown. Thus, the lack of a RL accuracy evaluation metric can lead to inaccuracy and inefficiency. To address this problem, this paper proposes a novel Dynamic Localization Graph (DLG) to both evaluate and enhance the RL accuracy. First, without prior knowledge of the relative poses, we propose a method for obtaining the Metric of Relative Localization Accuracy (MRLA) based on the pose covariance. Graduated non-convexity factor graph optimization is applied to calculate the transformation and update the pose covariance between the initial frames, followed by outlier rejection through hypothesis testing. Then, the DLG is constructed by assigning the MRLA between robots as the weight of each edge, which can clearly illustrate the status of RL in the multi-robot system and offer guidance for subsequent localization enhancement. The DLG is independent of the sensor system utilized. A place recognition-based RL method combined with the DLG is given as a case study. Simulations demonstrate that when the MRLA in the DLG reaches 0.98, the Absolute Pose Error is less than 0.1. Furthermore, the effectiveness of the DLG is confirmed through indoor and outdoor experiments. Project Page: <http://111>.

Index Terms—Multi-robot systems, relative localization, dynamic localization graph

I. INTRODUCTION

RELATIVE Localization (RL) enables multiple robots to transform their states from local to global frames, which is essential for tasks such as collaborative exploration [1] and surveillance [2]. Although GPS can provide high-accuracy localization outdoors, it is usually unavailable in GPS-denied environments such as indoor settings, which are common in robotic applications. To address this limitation,

Manuscript received Month xx, 2xxx; revised Month xx, xxxx; accepted Month x, xxxx. This work was supported in part by the National Natural Science Foundation of China (U1813224) and in part by the MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, China. (Corresponding author: Zhenhua Xiong.)

The authors are with the State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: cgm1015@sjtu.edu.cn; coldtea@sjtu.edu.cn; liuwenhang@sjtu.edu.cn; mawenyao@sjtu.edu.cn; mexiong@sjtu.edu.cn).

recent research has explored RL methods relying on onboard sensors, including UWB [3], Lidar [4], and Cameras [5].

In most multi-robot exploration tasks [6], once a single estimation between two robots is obtained, a determined relative pose is considered to be acquired. Since onboard sensors inevitably introduce errors, the relative pose estimation method also contains errors, which will propagate the local inaccuracy into the global frame. Thus, the current approach overlooks the uncertainty during RL, which may cause degradation in the merged global information, and potentially lead to erroneous decisions and failures in collaborative tasks. Thus, RL with single-instance pose estimation can be error prone.

Considering the imperfection of estimated relative poses, some researchers aim to minimize pose errors through multiple estimations. For example, in multi-robot simultaneous localization and mapping (SLAM), where place recognition is often utilized for RL, most strategies involve one robot repeating the trajectories of others [7]–[9] to obtain co-observation and optimize poses for better accuracy. However, existing studies mainly utilize the trajectory repetition empirically and lack the accuracy evaluation of RL results. This deficiency may cause significant inefficiency in some multi-robot applications. In multi-robot active SLAM, fast map reconstruction and accurate RL are usually in conflict, which is known as the exploration-exploitation dilemma [10]. Accurate RL favors re-visiting previously explored areas, while fast map reconstruction demands less trajectory repetition. Thus, a quantitative evaluation of RL results is needed to determine whether a robot should continue trajectory repetition or start a new exploration.

Therefore, a novel Dynamic Localization Graph (DLG) is proposed to achieve efficient relative localization through RL accuracy evaluation. Each robot constitutes a node of the DLG, and a connected graph is established between robots that have obtained RL results. Unlike previous methods that only use estimated relative poses, the DLG stores the optimized approximate posterior covariance between each robot's initial frame at the node and utilizes the Metric of Relative Localization Accuracy (MRLA) as the weight of each edge based on this covariance. The DLG is a general framework that is independent of the sensor system and can be combined with various RL methods. It can determine which robots have inaccurate RL results and coordinate robots to actively enhance localization. In Section III, the general framework of the DLG is established. It takes the odometry and several parameters as inputs, including pose estimation value, covariance, and confidence of each RL event, and calculates the relative poses and MRLA between the initial frames. First,

graduated non-convexity (GNC) [11] factor graph optimization and uncertainty propagation are performed on multiple RL events to obtain the approximate posterior distribution of the initial frames. Then, outliers are further eliminated through hypothesis testing, and the MRLA is calculated and updated based on the pose covariance, which is utilized to construct the DLG. Finally, RL enhancement is achieved by considering both the MRLA and efficiency. A specific case study of place recognition-based RL combined with the DLG is given, along with methods for computing the input parameters in Section IV. The main contributions are:

- 1) A quantitative evaluation method for RL accuracy is proposed based on the optimized posterior covariance between the initial frames of robots.
- 2) The DLG, a general framework and utilizes the MRLA as the weight of each edge, is proposed to achieve efficient relative localization by considering the uncertainty of RL and the enhancement cost.
- 3) A place recognition-based RL method combined with the DLG is given as a case study, and the detailed parameter computation methods are introduced. Extensive simulations and experiments demonstrate that when the MRLA reaches 0.98, the Absolute Pose Error is less than 0.1.

II. RELATED WORK

A. Relative Localization

Some researchers focus on single-instance RL, aiming to enhance the stability and accuracy of recognition, which is a foundation for multi-instance RL. A hybrid descriptor combining semantic and geometric information was proposed in [12], addressing challenges of viewpoint changes and seasonal variations. In [13], spherical harmonics-based viewpoint-invariant feature learning and a coarse-to-fine fast sequence-matching mechanismis are proposed to achieve robust place recognition. Despite efforts to predict the robot's pose using a convolutional neural network (CNN), outliers persist in [14]. As robots move, multiple RL events occur, prompting research into data association and optimization in multi-instance RL. The authors introduced an adaptive loop closure detection module for large-scale environments [4]. In [7], [15], outliers were handled through GNC pose graph optimization. In [9], spectral analysis was utilized to detect errors in pose graphs. In [16], considering aerial drag effects, an optimized key-node selection method is proposed to enhance localization accuracy. Our previous work [17] proposed a hierarchical RL framework whose high accuracy relies on pose graph optimization with multiple estimations.

Despite that significant progress has been made by the aforementioned methods, they either rely solely on a single result or depend on multiple localization optimizations empirically. However, uncertainty information of pose estimation has not been used to assess whether high-accuracy RL is achieved.

B. Multi-Robot Active Localization

After single-instance RL, multiple robots need to perform target selection and task allocation to actively enhance localization. The significance of multi-robot cooperative active

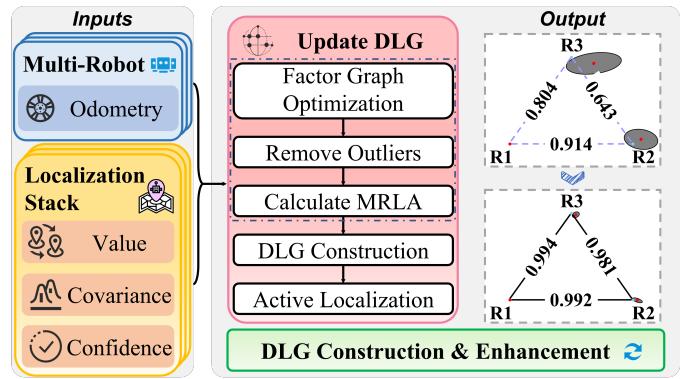


Fig. 1. Overview of the proposed architecture. The DLG is constructed and enhanced using the odometry and relative localization stack results of multiple robots.

estimation was highlighted in [10]. In [18], the graph topology metrics are used to evaluate the estimation uncertainty in the pose graph optimization. [19] further extended it to multiple robots and proposed an efficient two-stage strategy to achieve graph exploration. In [15], a loop closure candidate prioritization method was provided based on algebraic connectivity maximization. In [20], when the robot detected overlapping trajectories, it entered an adaptive merge state for pose validation and resumed exploration tasks after reaching a threshold. In [21], the robot moved to the adjacent topological node and compared the map structures to remove erroneous results. In [22], indirect belief constraints on multiple robots in future observations were introduced in belief space planning. In [23], an active persistent localization scheme that integrates an enhanced set-membership filter for cooperative data fusion and motion planning under optimal localization conditions is proposed. However, it focuses on the collaborative localization of the target and ignores the RL uncertainty between robots.

Existing multi-robot active localization methods concentrate on selecting localization points and planning path repetition strategies to promote RL accuracy. However, since the ground truth is unknown in real applications, there has been limited research on RL accuracy evaluation and multi-robot coordination for efficient relative localization.

III. METHODOLOGY

The framework of the DLG construction and enhancement is shown in Fig. 1, where odometry and RL information from multiple robots are as inputs, and relative poses and accuracy between initial frames are acquired. Confidence signifies the reliability of each pose estimation event, and the trace of the pose covariance denote its precision. To evaluate the accuracy of RL, the robot odometry, along with the value and measurement covariance of each pose estimation event are utilized for factor graph optimization to obtain the optimized pose and posterior pose covariance between initial frames. Erroneous results are discarded using hypothesis testing, and the pose covariance is utilized to calculate the MRLA. The DLG, using the MRLA as edge weights, is constructed and updated incrementally. Finally, based on the DLG, target selection and task allocation for RL enhancement are performed.

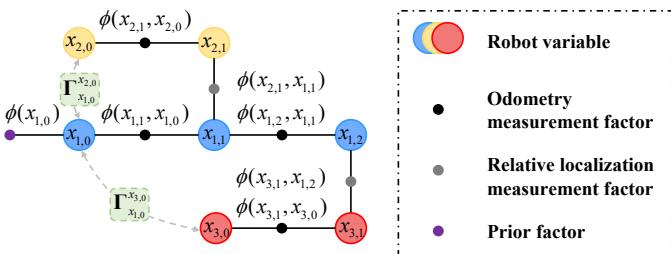


Fig. 2. Factor graph instance, which mainly contains robot state variables, odometry measurement factors and relative localization measurement factors.

A. Evaluation of Relative Localization Accuracy

When the localization stack has multiple RL events, factor graph optimization can minimize the pose error, and the distribution of each robot's initial frame is obtained through uncertainty propagation. Considering the uncertainty of the pose estimation method, wrong results are further eliminated by hypothesis testing. Finally, the MRLA is calculated based on the optimized pose covariance.

1) *Factor Graph Optimization with GNC*: The states and observations of multiple robots are represented as a factor graph, where a basic example is illustrated in Fig. 2. State variables of different robots are connected through RL measurement factors. The adjacent variable is defined to symbolize the set of pose variables neighboring factor ϕ . For ease of representation, any adjacent variables connected by a measurement factor in Fig. 2 are denoted as (x_a, x_b) . The optimization problem is presented as follows:

$$\mathbf{X} = \arg \max_{\mathbf{X} \in SE(3)} \prod_{(x_a, x_b) \in O(\mathbf{X})} \phi(x_a, x_b) \prod_{(x_a, x_b) \in L(\mathbf{X})} \phi(x_a, x_b) \quad (1)$$

where \mathbf{X} is the set of state variables for robots, $O(\mathbf{X})$ represents the set of adjacent variables for all odometry and prior factors, and $L(\mathbf{X})$ is the set of adjacent variables for all RL measurement factors.

For each factor, the residual vector is defined as follows:

$$\mathbf{r}(\mathbf{T}_{x_b}^{x_a}) = \left\| \mathbf{T}_{x_b}^{x_a} \boxminus \hat{\mathbf{T}}_{x_b}^{x_a} \right\|_{\Sigma_{x_b}^{x_a}} \quad (2)$$

where $\Sigma_{x_b}^{x_a}$ is the measurement covariance of the relative pose estimation or odometry, and \boxminus denotes a tangent space representation of the relative pose between $\mathbf{T}_{x_b}^{x_a}$ and $\hat{\mathbf{T}}_{x_b}^{x_a}$ [7].

Generally, odometry is more accurate than RL, which sometimes consist of outliers. To reduce the interference of outliers caused by erroneous data association, low-confidence and high-variance RL results are directly discarded. For the remaining RL factors, a robust objective function, denoted as Ψ , is employed. Then, Eq. (1) can be transformed as follows:

$$\min_{\mathbf{X} \in SE(3)} \sum_{(x_a, x_b) \in O(\mathbf{X})} \mathbf{r}^2(\mathbf{T}_{x_b}^{x_a}) + \sum_{(x_a, x_b) \in L(\mathbf{X})} \Psi(\mathbf{r}(\mathbf{T}_{x_b}^{x_a})) \quad (3)$$

For $\Psi(\mathbf{r}(\mathbf{T}_{x_b}^{x_a}))$, according to Black-Rangarajan Duality [24], the robust non-linear least squares problem (Eq. (3)) is

equivalent to the following decoupled formulation:

$$\begin{aligned} \min_{\mathbf{X} \in SE(3), \mathcal{W} \in [0, 1]} & \sum_{(x_a, x_b) \in O(\mathbf{X})} \mathbf{r}^2(\mathbf{T}_{x_b}^{x_a}) + \\ & \sum_{(x_a, x_b) \in L(\mathbf{X})} \omega_{ab} \mathbf{r}^2(\mathbf{T}_{x_b}^{x_a}) + \Phi_\rho(\omega_{ab}) \end{aligned} \quad (4)$$

where ω_{ab} is the weight of the residual of adjacent variables associated with the RL factor, \mathcal{W} denotes the set of weights, and $\Phi_\rho(\omega_{ab})$ is the penalty function. In the implementation, the Truncated Least Squares (TLS) function is used. To solve the challenges of non-convex optimization, GNC utilizes a series of surrogate functions to approximate the non-convex optimization problem [25]. It achieves this process by updating variables, weights, and parameters [11]. The GNC implementation from the GTSAM library [26] is employed, utilizing the Levenberg-Marquardt algorithm (LMA) for variable updates.

The pose covariance of adjacent variables associated with measurement factors is calculated, which is determined by the corresponding measurement covariance, factor graph structure and optimized pose. At the k -th iteration of the LMA, the linearization approximation and Laplace approximation are adopted, and the pose covariance is adjusted following:

$$\mathbf{k+1}\Gamma_{x_b}^{x_a} = \begin{cases} \Sigma_{x_b}^{x_a}, & k = 0 \\ \left((\mathbf{J}^k)^\top (\mathbf{k}\Gamma_{x_b}^{x_a})^{-1} \mathbf{J}^k \right)^{-1}, & k \geq 1 \end{cases} \quad (5)$$

where $\Gamma_{x_b}^{x_a}$ denotes the pose covariance of the adjacent variables, and \mathbf{J}^k is the Jacobian at the current linearization point, corresponding to the variables of each measurement factor.

The relative transformation between the initial frames of robots consists of a compound pose change. To calculate the uncertainty in this process, [27] is utilized to derive the uncertainty propagation. The difference is that the right-hand convention is adopted, which is consistent with the definition of the GTSAM library. Without loss of generality, we provide the uncertainty propagation equation that characterizes $\mathbf{T}_{x_c}^{x_a} = \mathbf{T}_{x_b}^{x_a} \mathbf{T}_{x_c}^{x_b}$ and $\mathbf{T}_{x_d}^{x_a} = \mathbf{T}_{x_c}^{x_a} (\mathbf{T}_{x_c}^{x_d})^{-1}$ respectively:

$$\begin{aligned} \Gamma_{a,c} &\approx \text{Ad}_{\mathbf{T}_{b,c}^{-1}} \Gamma_{a,b} \text{Ad}_{\mathbf{T}_{b,c}^{-1}}^\top + \Gamma_{b,c} + \\ &\text{Ad}_{\mathbf{T}_{b,c}^{-1}} \Gamma_{ab,bc} + \Gamma_{ab,bc}^\top \text{Ad}_{\mathbf{T}_{b,c}^{-1}} \end{aligned} \quad (6)$$

$$\Gamma_{a,d} \approx \text{Ad}_{\mathbf{T}_{d,c}} (\Gamma_{a,c} + \Gamma_{d,c} - \Gamma_{ac,dc} - \Gamma_{ac,dc}^\top) \text{Ad}_{\mathbf{T}_{d,c}}^\top \quad (7)$$

Note that for convenience of notation, superscripts are written in subscripts in Eq. (6) and (7), expressing the same meaning. Ad_T is the matrix form of the adjoint action of T . All inputs are obtained through factor graph optimization.

Assume the initial frame of robot 1 serves as the global reference frame. The initial poses of the remaining robots are set using the identity matrix in the beginning. Following the optimization mentioned earlier, the relative pose distribution of the initial frame between robots r and s , interconnected through RL, can be obtained as $\eta_{x_s,0}^{x_r,0} \sim \text{Gaussian}(\bar{\eta}_{x_s,0}^{x_r,0}, \Gamma_{x_s,0}^{x_r,0})$.

2) *Outlier Removal*: Outlier detection is performed utilizing hypothesis testing for the low-confidence and low-variance RL result that has a high precision. The relative pose of the initial frames between robots r and s is calculated with Eq. (8). Due to the correlation between the degrees of freedom (DoFs)

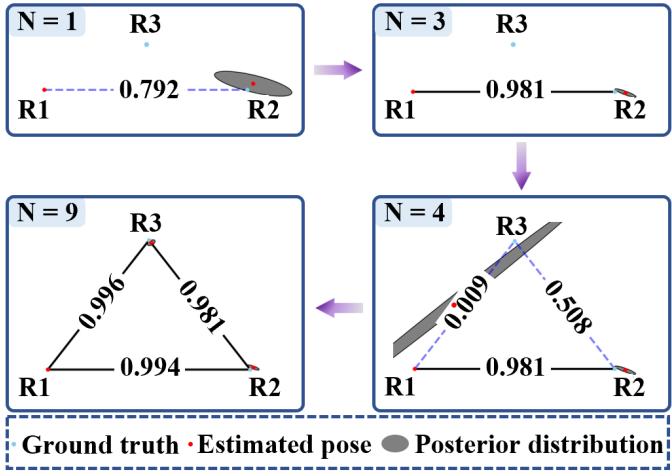


Fig. 3. As the number of RL events increases, the DLG is dynamically constructed and updated. Ultimately, the uncertainty of each robot's initial frame is reduced and converges to small ellipses surrounding the ground truth.

in the robot pose, the Mahalanobis Distance between the current result $\hat{\eta}_{x_s,0}^{x_r,0}$ and the optimized result $\bar{\eta}_{x_s,0}^{x_r,0}$ is calculated as the statistic parameter with Eq. (9).

$$\hat{\eta}_{x_s,0}^{x_r,0} = \log \left(\hat{\mathbf{T}}_{x_r,i}^{x_r,0} \hat{\mathbf{T}}_{x_s,j}^{x_r,i} (\hat{\mathbf{T}}_{x_s,j}^{x_r,0})^{-1} \right) \quad (8)$$

$$d_M = \sqrt{(\hat{\eta}_{x_s,0}^{x_r,0} - \bar{\eta}_{x_s,0}^{x_r,0})^\top (\mathbf{\Gamma}_{x_s,0}^{x_r,0})^{-1} (\hat{\eta}_{x_s,0}^{x_r,0} - \bar{\eta}_{x_s,0}^{x_r,0})} \quad (9)$$

The squared distance d_M^2 is compared with $\chi^2(6)$ based on the chi-square distribution approximation to assess if this result falls within the distribution. The null hypothesis (H_0) is established, stating that $\hat{\eta}_{x_s,0}^{x_r,0}$ is within $Gaussian(\bar{\eta}_{x_s,0}^{x_r,0}, \mathbf{\Gamma}_{x_s,0}^{x_r,0})$. Otherwise, the alternative hypothesis (H_1) is considered. At a 95% level, $\chi_{0.95}^2(6)$ is chosen as the critical value. If $d_M^2 > \chi_{0.95}^2(6)$, H_0 is rejected, indicating that the current RL result is an outlier. As a result, it is removed from the localization stack.

3) *Accuracy Evaluation using the MRLA:* The covariance is utilized to measure the precision of the RL between any two robots. Upon updating the distribution of initial frames, the dispersion of all DoFs is determined by the trace of the pose covariance, represented as $\text{trace}(\mathbf{\Gamma}_{x_s,0}^{x_r,0})$. Then, the accuracy of the initial frames between robots r and s is calculated as the MRLA following:

$$e_{r,s} = \exp \left(-\frac{\text{trace}(\mathbf{\Gamma}_{x_s,0}^{x_r,0})}{6} \right) \quad (10)$$

The MRLA is similar to A-optimality [28], and we use a negative exponential mapping to transform it into the interval $(0, 1)$ for easier quantification. Thus, without prior knowledge of the relative poses, the RL accuracy between robots can be determined through the above evaluation.

B. DLG Construction

After calculating the MRLA of the robot swarm that has been relatively localization, the DLG is constructed to clearly illustrate the RL accuracy of multi-robot systems and offer

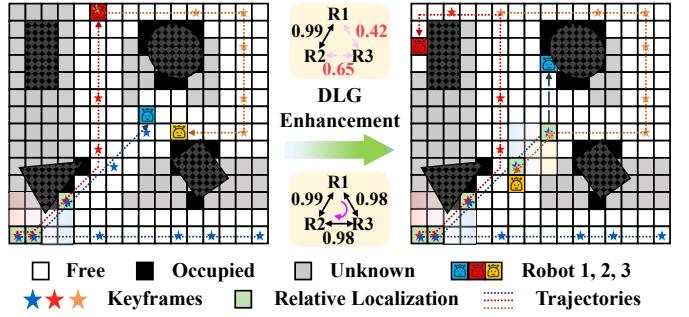


Fig. 4. The process of multi-robot active localization enhancement. Based on the comprehensive consideration of RL accuracy and cost, robot 3 is chosen to repeat the trajectory of robot 1.

guidance for subsequent localization enhancement. The topology of the DLG is represented by a connected graph consisting of all robot nodes that has already obtained RL results.

Fig. 3 illustrates the process of DLG construction with active localization enhancement (Section III-C), where each node represents the robot's identifier, and the MRLA is set as the weight of each edge that reflects the RL accuracy between two robots. To intuitively demonstrate the evaluation effect of the MRLA, the figure also displays the ground truth and the approximate posterior distribution of the initial frame of each robot in the global reference frame, with the initial frame of robot 1 used as a reference. A different translation transformation is applied to both for easier comparison. In the DLG, the ground truth is solely used to indicate the distance between it and the estimated result, without being involved in any calculations. As the number of RL events increases, the trace of the robot's initial frame's pose covariance decreases, MRLA approaches 1, and the estimated pose progressively converges to the ground truth.

C. Active Localization Enhancement

After obtaining preliminary RL results, it is imperative to coordinate the robot swarm to make individual decisions on whether to continue their primary task or engage in RL enhancement. When robot t joins the DLG, if $e_{1,t}$ is less than a given threshold λ , an active localization enhancement approach is needed. To achieve this goal, the following optimization is performed:

$$\mathcal{R}_M, G_M = \arg \min_{P \in Path(1,t)} L_P \sum_{r,s \in P} e_{r,s}^{-1} C(r,s) \quad (11)$$

where $Path(1,t)$ is the set of all reachable paths from robot 1 to t in the DLG, L_P denotes the number of robot nodes in the current path P , r and s are two adjacent robot nodes in this path, and $C(r,s)$ is the cost required for robots r and s to perform localization enhancement. Eq. (11) is based on the consideration of minimizing accumulative errors and costs, while making the best use of the relatively accurate RL results already available. As a result of this optimization process, a set of robots \mathcal{R}_M and a set of target points G_M are identified for RL enhancement. If $e_{r,s} > \lambda$, it indicates that accurate RL has already been achieved between robots r and s , in which case the corresponding elements are removed

Algorithm 1: DLG Construction and Enhancement

Input: Odometry set \mathcal{O} ; Relative localization stack \mathcal{L} ;
Accuracy threshold λ

Output: Optimized initial frames' relative poses \mathbf{P}
and pose covariance Γ ; Enhanced DLG \mathbf{G}

```

1 while Robots in Relative Localization do
2    $\mathbf{P}, \Gamma \leftarrow \text{GNCFG}(\mathcal{O}, \mathcal{L});$ 
3   for  $l$  in  $\mathcal{L}$  do
4     |  $\mathcal{L} \leftarrow \text{OutlierRemoval}(\mathbf{P}, \Gamma, \mathcal{O}, l);$ 
5   end
6    $e \leftarrow \text{MRLACalculation}(\Gamma);$ 
7    $\mathbf{G} \leftarrow \text{DLGConstruction}(\mathbf{P}, \Gamma, e);$ 
8   if  $\exists e_{1,t} \leq \lambda, e_{1,t} \in \mathbf{G}$  then
9     |  $\text{LocalEnhance}(\mathbf{G}, \mathcal{O}, \mathbf{P});$ 
10  end
11 end
12 return  $\mathbf{P}, \Gamma, \mathbf{G}$ 
```

from \mathcal{R}_M and G_M . Therefore, by considering the MRLA and enhancement cost for all robots comprehensively, this strategy improves the RL accuracy between robots 1 and t . The acquisition of \mathcal{R}_M and G_M depends on the defined cost function, which varies depending on the specific RL method. We provide a computational example of the place recognition-based RL method in Section IV.

An example is shown in Fig. 4, where RL is achieved at keyframes. Upon the initial construction of the DLG, the MRLA between robots 1 and 3, as well as between 2 and 3, is low. Robot 2 refrains from repeating robot 3's trajectory, given that it is farther from robot 3's subsequent keyframe. On the contrary, robot 3 actively replicates the trajectory of robot 1, because it is closer to robot 1's keyframe and has a smaller accumulative error when directly transformed to the global frame corresponding to the initial frame of robot 1. Consequently, both robot 1 and 2 continue their primary tasks, while robot 3 actively repeats robot 1's trajectory to achieve RL enhancement.

Overall, algorithm 1 is employed to dynamically maintain and update the DLG, and achieve efficient relative localization through relative localization accuracy evaluation, the DLG construction, and active localization enhancement.

IV. CASE STUDY WITH PLACE RECOGNITION

Since the DLG is a general framework, in order to demonstrate its specific application, the architecture of a place recognition-based RL method combined with the DLG is depicted in Fig. 5. Algorithm 2 illustrates the workflow. Due to the effectiveness of our previous work [29], keyframes are set to store the pose, pointclouds, and image descriptor output by a CNN [30]. Real-time descriptor comparison with keyframes determines if the robot has encountered a similar place. The relative pose is then obtained by performing a global Iterative Closest Point (ICP) algorithm with the captured pointclouds and the candidate keyframe pointclouds. Subsequently, the DLG is utilized to optimize and assess the RL result, using the calculated confidence and covariance. This process identifies

the robots requiring RL enhancement and determines their targets. The calculation of the parameters required by Algorithm 1 is given below:

1) *Confidence Calculation:* To determine if the image descriptor \mathbf{D}_r from robot r has a similarity in the keyframe descriptor stack \mathcal{D} established by all robots, the cosine similarity metric is employed. The threshold is set to φ_{th} . If it exists, denote the image descriptor with the highest similarity as \mathbf{D}_s . To reduce the impact of false positives in place recognition, the corresponding image descriptors from all similar keyframes are extracted to form the set \mathcal{S}_r .

$$\mathcal{S}_r = \{\mathbf{D}_t | \mathbf{D}_t \in \mathcal{D}, \cos(\mathbf{D}_r, \mathbf{D}_t) > \tau \cos(\mathbf{D}_r, \mathbf{D}_s)\} \quad (12)$$

When calculating the confidence C_s for each RL event, both the specificity and diversity of the descriptors are considered. The distance factor α_t is introduced to normalize the distances between descriptors in \mathcal{S}_r to characterize the spatial distribution differences of keyframes with similar features. Considering both the similarity and the distance, C_s is calculated to represent the reliability of each RL event as follows:

$$C_s = \frac{1}{|\mathcal{S}_r|} \sum_{\mathbf{D}_t \in \mathcal{S}_r} [\cos(\mathbf{D}_r, \mathbf{D}_s) - \alpha_t \cos(\mathbf{D}_r, \mathbf{D}_t)] \quad (13)$$

$$\alpha_t = \begin{cases} \frac{\|\mathbf{p}_t - \mathbf{p}_s\|_2}{\max_{\mathbf{D}_u \in \mathcal{S}_r, e_{s,u} \neq 0} \|\mathbf{p}_u - \mathbf{p}_s\|_2}, |\mathcal{S}_r| > 1, e_{s,t} \neq 0 \\ 1, |\mathcal{S}_r| > 1, e_{s,t} = 0 \\ 0, |\mathcal{S}_r| = 1 \end{cases} \quad (14)$$

where \mathbf{p}_t and \mathbf{p}_s represent the position of the keyframe in the global reference frame.

2) *Value and Covariance Calculation:* After place recognition, Global ICP is utilized for RL. The process begins with the global registration using RANSAC upon pointcloud features. Subsequently, local registration is performed using Stein ICP [31] to calculate the relative pose and model the measurement uncertainty because it is efficient and accurate. A set of particles is randomly initialized with the initial solution. Stein variational gradient descent is applied to update each particle, iterating multiple times to obtain the distribution of the particles, thereby acquiring the pose estimation value $\mathbf{T}_{x_{s,j}}^{x_{r,i}}$ and measurement covariance $\Sigma_{x_{s,j}}^{x_{r,i}}$ of each RL event.

Through the aforementioned calculations, all inputs required for the DLG, including the odometry, value, covariance and confidence of each RL event are obtained. Subsequently, global optimization provides the approximate posterior probability distribution between the initial frames of the robots that has been relatively localized.

3) *Cost Definition:* When enhancing RL, taking into account the robot's motion distance and consistency, the cost function $C(r, s)$ is defined as the minimum movement cost for robot r to transition from its current pose \mathbf{p}_r to the pose $\mathbf{p}_{s,k}$ that belongs to a keyframe k that has not yet participated in RL of robot s :

$$C(r, s) = \min_{k \in \mathcal{K}_s} \frac{\text{dis}(\mathbf{p}_r, \mathbf{p}_{s,k})}{v_m} + \frac{\theta_k}{\omega_m} \quad (15)$$

$$\theta_k = \arccos \frac{(\mathbf{p}_{s,k} - \mathbf{p}_r) \cdot \mathbf{p}_r}{\|\mathbf{p}_{s,k} - \mathbf{p}_r\|_2 \cdot \|\mathbf{p}_r\|_2} \quad (16)$$

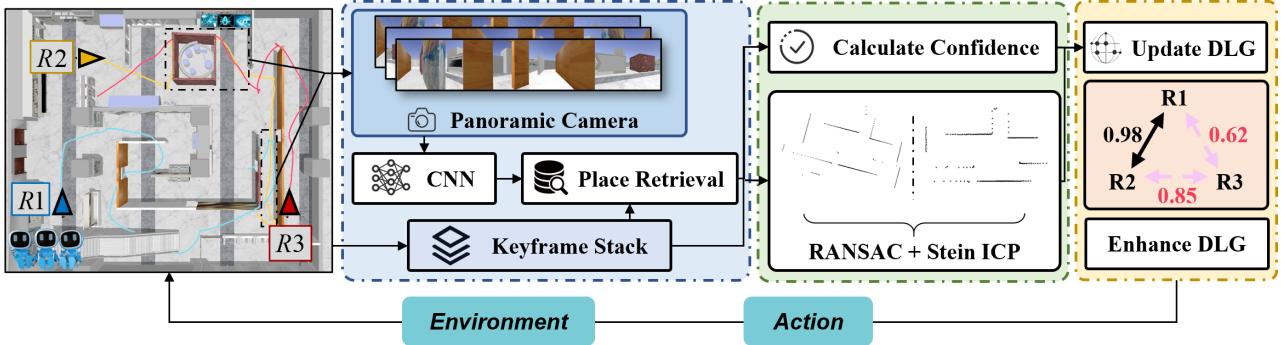


Fig. 5. The architecture of integrating the DLG with a place recognition-based RL method in a multi-robot system. Firstly, the robots set keyframes in the environment. Then, panoramic images are captured to extract descriptors for matching, and confidence is calculated. Subsequently, the pointclouds is utilized to compute the relative pose and covariance. Finally, above information is used to guide the DLG update and enhancement.

Algorithm 2: Place Recognition with the DLG

```

Input:
- Initial DLG  $G$ ; Keyframe stack  $\mathcal{K}$ ;
- Odometry set  $O$ ; Image set  $I$ ; Pointcloud set  $L$ ;
- Initial relative localization stack  $\mathcal{L}$ 
Output:
- Optimized initial frames' relative poses  $P$  and
  pose covariance  $\Gamma$ ; Enhanced DLG  $G$ 
1 while Robots in Motion do
2   for  $r$  in  $\mathcal{R}$  do
3      $D_r \leftarrow \text{CNN}(I_r)$ ;
4      $D_s, S_r \leftarrow \text{FindSimilarity}(D_r, \mathcal{K})$ ;
5     if  $S_r \neq \emptyset$  then
6        $C_s \leftarrow \text{CalcConf}(D_r, D_s, S_r, G)$ ;
7        $T_{x_s, j}^{x_r, i}, \Sigma_{x_s, j}^{x_r, i} \leftarrow \text{Stein-ICP}(L_r, \mathcal{K})$ ;
8        $\mathcal{L} \leftarrow \text{Insert}(T_{x_s, j}^{x_r, i}, \Sigma_{x_s, j}^{x_r, i}, C_s)$ ;
9        $G, P, \Gamma \leftarrow \text{DLGConstEnhance}(O, \mathcal{L})$ ;
10      end
11    end
12  end
13 return  $P, \Gamma, G$ 

```

where \mathcal{K}_s represents the keyframe stack of robot s , $dis()$ calculates the Euclidean distance between the two points, θ_k is the required turning angle to move to the target, and v_m and ω_m are the linear and angular velocities, respectively.

The cost of RL enhancement between any two robots in the DLG and the corresponding target of each robot are calculated using Eq. (15). Substituting them into Eq. (11), the cost of all feasible solutions is calculated utilizing the enumeration algorithm, and finally the solution with the minimum cost is selected to construct \mathcal{R}_M and G_M .

V. SIMULATIONS AND EXPERIMENTS

A. Experimental Setup

Simulations and experiments are conducted using the Turtlebot3 Burger robot, equipped with a 2D Lidar and a 360-degree panoramic camera. The hardware platform consists of an Intel i9 3.2GHz CPU and a GTX 3080 Ti GPU. All experiments are based on place recognition for RL, employing

TABLE I
LOCALIZATION RESULTS WITH DIFFERENT NUMBERS OF ROBOTS

Robot Num	Mapping Time (s)	Map Accuracy (%)	APE
2	186.1	96.5	0.082
3	129.8	96.6	0.053
4	119.4	95.7	0.093

the method detailed in Section IV. In the implementation, the accuracy threshold λ for localization enhancement is 0.98 to balance the accuracy and efficiency, indicating that accurate RL is achieved when the MRLA exceeds 0.98 between two robots. The similarity threshold φ_{th} for place recognition is 0.97 [29], and the weight τ for forming a set of similar descriptors is 0.97 to balance the impact of false positives and computational effort. The Turtlebot3 Burger's maximum linear velocity v_m is 0.22 m/s, and the maximum angular velocity ω_m is 2.84 rad/s. Due to our focus on measuring the uncertainty of RL between robots, the robot's odometry is obtained using the SLAM algorithms to be more accurate. To illustrate the uncertainty of the robot's self-localization, we set the pose variance between adjacent keyframes to a constant value. Due to limited onboard computing resources, all robot data is transmitted to the computer, and the system operates in a centralized manner.

B. Case Study

To verify the evaluation and planning performance of the DLG in multi-robot systems, case studies involving different numbers of robots are conducted in the Gazebo software. The simulation area covers 485 m². Each robot utilizes Cartographer for SLAM to obtain the odometry, with map accuracy and Absolute Pose Error (APE) serving as evaluation metrics. Map accuracy is measured by the proportion of correct cells between the aligned ground-truth map and the merged map constructed by multiple robots. APE is calculated according to Eq. (17), where N_t represents the number of keyframes of all robots, and $T_{gt,i}$ and $T_{est,i}$ denote the ground truth and estimated localization result in the global frame, respectively.

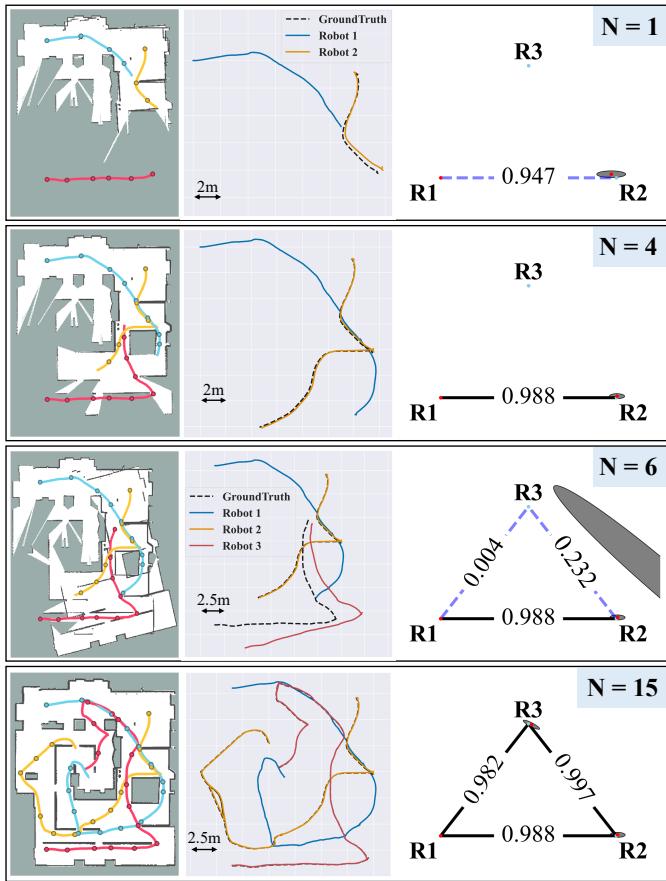


Fig. 6. The process of RL optimization and assessment of three robots based on the DLG, including the merged map, robot trajectories, and the DLG, etc. The robots achieve RL at keyframes (indicated by dots on the trajectories). The results of the first, fourth, sixth, and 15th times are shown respectively.

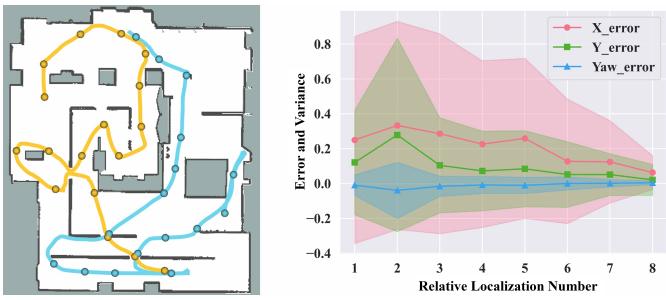


Fig. 7. The results of RL for two robots with the DLG.

The results are presented in Table I.

$$\text{APE} = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} \left\| \log(T_{gt,i}^{-1} T_{est,i})^{\top} \right\|_2^2} \quad (17)$$

Fig. 6 demonstrates the process of three robots mapping and optimizing RL. Robot 1 and 2 initially establish a comparatively accurate relative pose, achieving the MRLA of 0.988 after three active localizations, meeting the stopping criteria for localization enhancement. At this point, the optimized robot trajectories closely align with the ground truth. Due to the sparse pointcloud features in the first RL area for Robot 3,

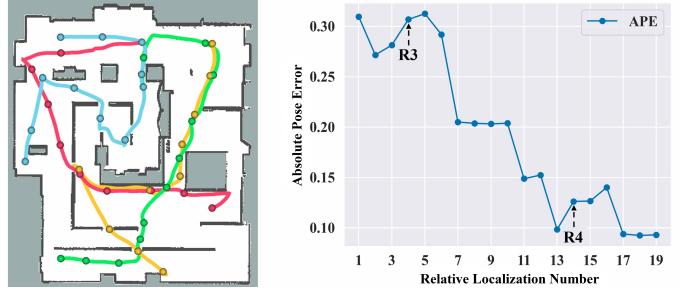


Fig. 8. The results of RL for four robots with the DLG.

TABLE II
LOCALIZATION RESULTS OF DIFFERENT METHODS WITH THREE ROBOTS

Methods	Mapping Time (s)	Map Accuracy (%)	APE	Num
SL	109.1	81.5	0.453	2
FV	148.8	94.1	0.098	33
LR	189.6	96.6	0.050	45
CGE [19]	141.3	95.9	0.064	26
Ours	129.8	96.6	0.053	15

the initial localizations result in significant errors and variance. Subsequent active localizations, in which Robot 3 replicates Robot 1's trajectory, help to optimize the pose and reduce the offset in the global frame. Finally, after a total of 15 RL events, a stable DLG is constructed. The scene reconstruction takes 129.8 seconds, reducing the APE to 0.053.

Fig. 7 illustrates the RL results of two robots, including their respective trajectories and the changes of errors and variances of each DoF. Following two RL events, both errors and variances of each DoF exhibit a decreasing trend. Eventually, the termination condition is reached after 8 RL events, and the variances and errors converge near the origin. The APE decreases to 0.082, achieving a mapping accuracy of 96.5%. Fig. 8 presents the results of four robots and the trend of APE change as the number of RL event increases. Initially, Robot 1 and 2 implement RL and reduce the overall trajectory error with active localization enhancement. The participation of Robot 3 in the fourth RL causes an increase in APE. Then, as the number of active localizations increases, APE gradually decreases. After 13 RL events, a stable DLG for three robots is achieved. Robot 4 joins subsequently, and after 19 RL events, reliable RL among four robots is established. The mapping time only slightly decreases by 10.4 seconds compared to the three-robot setup because more trajectories need to be repeated. The final APE is 0.093.

C. Comparison With Other Methods

Taking the above three-robot setup as an example, our method is compared with the following four methods.

- Single Localization (**SL**): This approach considers only the pose estimation result of the first RL event.
- Fixed Variance (**FV**): This method uses fixed variance parameters, as specified in [7], with a translation variance

of 0.5 m and a rotation variance of 0.1 rad. The rest of this approach is consistent with our method.

- **Long Repetition (LR):** Robots empirically repeat a long segment of trajectories between each other. We set the existence of 15 RL events between each pair of robots as the termination condition of RL. Other aspects remain consistent with ours.
- **CGE [19]:** The graph is constructed and updated using topological nodes of the free space [29] and frontiers.

Table II presents the results of the five methods. Due to relying solely on a single RL event, **SL** exhibits the shortest mapping time but with the highest APE of 0.453, because the error of a single RL is relatively large. In **FV**, adopting the same covariance for RL events with different uncertainties in reality means that correct and incorrect results impose the same constraints on the pose graph, which affects the accuracy. The large variance necessitates 33 RL events to construct a stable DLG, leading to an extended mapping time. In **LR**, owing to the lack of evaluation of RL accuracy, the trajectory overlap length can only be performed empirically. Due to the increased number of RL events, the mapping time increases by 59.8 seconds compared to **Ours**. However, the mapping accuracy remains nearly unchanged, and APE only shows a slight decrease. In **CGE**, since the graph is constructed incrementally, the robots tend to explore the environment in the early stages of the task. However, in the later stages, since the number of topological nodes in the explored area is much larger than the number of unexplored frontiers, the robots tend to over-exploit the environment, thereby reducing efficiency and increasing its mapping time consumption by 8.9% compared to **Ours**. Our method effectively balances RL accuracy and mapping time, ultimately achieving efficient RL and active scene reconstruction among the three robots in a shorter time. Therefore, the DLG serves as an effective evaluation for RL, reducing the number of RL events while maintaining a small error, thereby enhancing the performance of the overall task.

D. DLG With MARS Dataset

To validate the effectiveness of the DLG in evaluating RL accuracy in complex real-world scenarios, we conduct tests using the MARS dataset [32]. Each vehicle is equipped with a 3D Lidar and multiple cameras to capture pointclouds and images. Multiple vehicles traverse nearby areas on city streets. We select four scenes from the MARS dataset and present the MRLA calculated by the DLG and the merged maps in Fig. 9. We still adopt the RL method mentioned above because the focus is on the evaluation of RL accuracy, and other methods are also applicable as long as they provide the parameters required by the DLG.

For Scene 6, the two vehicles move consecutively along a straight street with significant trajectory overlap. The final RL accuracy is 0.99, and the merged map (Fig. 9(a)) exhibits minimal deviation. In Scene 7, the two vehicles have substantial trajectory overlap in the first half and then separate after an intersection, resulting in an accuracy of 0.99. The optimized initial poses produce a merged map (Fig. 9(b))

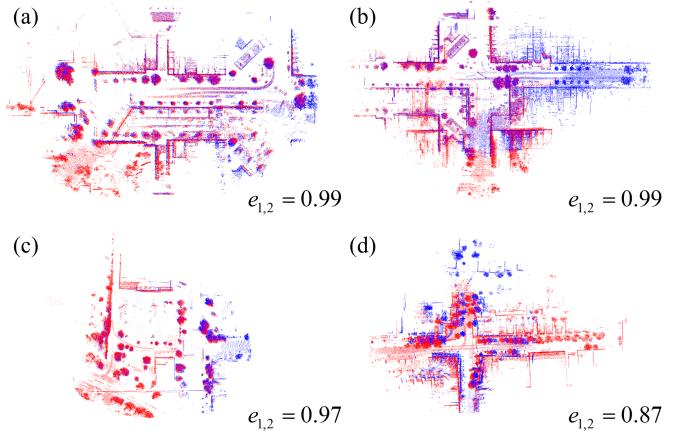


Fig. 9. The DLG is utilized to evaluate RL accuracy and merge map in typical real-world scenes from the MARS Dataset, where red pointclouds and blue pointclouds are collected by Agent 1 and Agent 2, respectively. (a) Scene 6. (b) Scene 7. (c) Scene 13. (d) Scene 27.

closely resembling the ground truth. In Scene 13, the vehicles share some overlap in the latter part of their motion. Due to the sparse features in the overlap segment, the variance calculated using Stein ICP remains large, yielding an accuracy of 0.97. The merged map (Fig. 9(c)) shows slight misalignment. For Scene 27, due to the short overlap segment, the two vehicles only achieve RL three times. The first two attempts show significant errors and variances due to the high similarity at the intersection. The final RL accuracy is only 0.87, and the merged map (Fig. 9(d)) shows a noticeable deviation. Thus, the DLG demonstrates applicability in real-world street scenarios, highlighting the results and accuracy of multi-agent RL.

E. Real-World Experiment

We also conduct real-world experiments involving two and three robots performing SLAM and RL combined with the DLG, respectively. The robot is equipped with a RPLidar for self-localization and mapping, and a RealSense D435i camera is utilized to capture images for place recognition.

For the case of two robots, the robot trajectories as well as the final merged map are depicted in Fig. 10(a). The robots enter the room independently from opposite sides of the corridor. Upon entering the room, Robot 2 detects the similar keyframe stored by Robot 1 near the door, thus achieving the first RL. Then, Robot 2 actively replicates the trajectory of Robot 1 to make RL more accurate. After completing RL and localization enhancement through 5 RL events, the MRLA of the two robots reaches 0.982. Subsequently, the two robots independently move in diverse directions to facilitate efficient reconstruction, achieving a map accuracy of 96.1%.

For the case of three robots, the robot trajectories, the final constructed DLG and the merged map are shown in Fig. 10(b). Initially, Robot 2 achieves RL with Robot 1 in front of the door and begins to repeat the trajectory of Robot 1. After achieving RL 4 times in total, the MRLA between the two robots reaches 0.983. Then, Robots 1 and 3, and Robots 2 and 3 achieve RL in the similar time. The robot swarm begin to coordinate enhanced localization, where Robot 1 first repeats

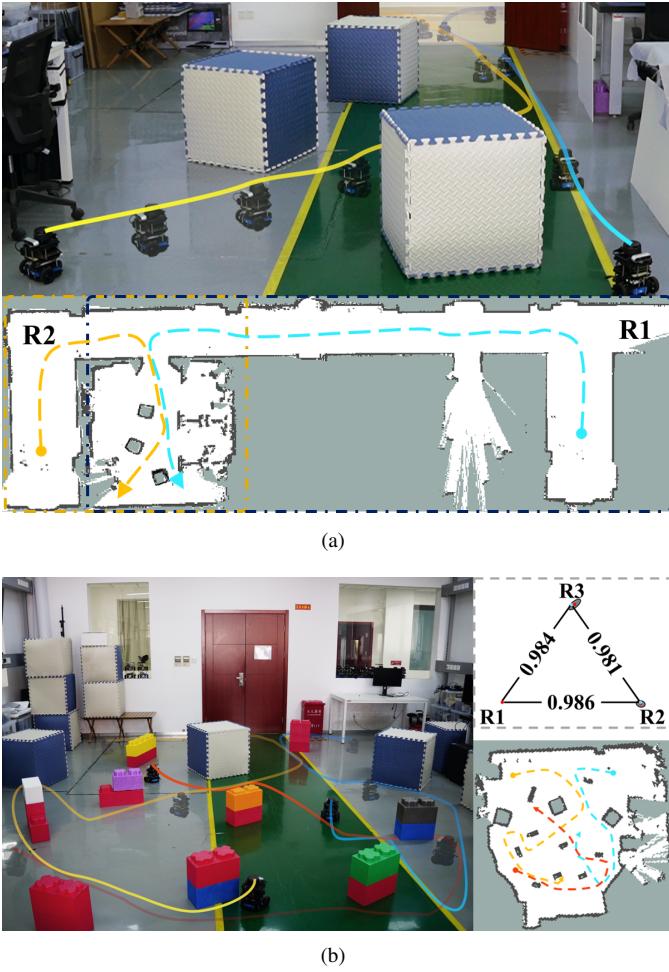


Fig. 10. Experimental results of two and three robots in real-world scenarios.

the trajectory of Robot 3, and the MRLA between the two robots reaches 0.963. Then, Robot 2 continues to repeat the trajectory of Robot 3, and Robot 3 explores the unknown area alone from beginning to end. Finally, after 19 RL events, all of the MRLA is greater than the set threshold, and the estimated robot's initial frame is close to the ground truth. The map accuracy under the three-robot setup reaches 95.4%.

F. Discussion

The DLG calculates optimized relative pose of multi-robot initial frames, along with an interpretable accuracy metric. Our work focuses on evaluating the uncertainty of the RL between multiple robots and rarely considers the localization of a single robot itself because its performance is already excellent [33]. It allows for setting suitable termination thresholds to balance the accuracy and the efficiency, denoted as λ , based on requirements. When active localization enhancement is not used, the DLG can be utilized for the assessment of RL accuracy, as shown in Section V-D. Therefore, the DLG can be employed together with the back-end of multi-robot SLAM to determine whether RL is still required and when to terminate trajectory repetition between robots. Furthermore, due to its principled abstraction, the DLG can be integrated

with other RL methods by computing the inputs from the sensor measurement model.

However, in highly repetitive environments, such as parking lots, the similar structure can lead to a large $|S_r|$. Consequently, the confidence of almost all observations becomes low, making it impossible to obtain the relative poses between robots. In such cases, multi-robot collaborative reconstruction becomes equivalent to single-robot reconstruction. A practical approach is to integrate rendezvous-based methods [17] with the aforementioned approach to establish a robust RL system.

VI. CONCLUSION

This paper proposes a novel Dynamic Localization Graph (DLG) to achieve efficient relative localization by considering the RL uncertainty and enhancement cost. GNC factor graph optimization and uncertainty propagation are used to calculate the RL results and pose covariance, and then hypothesis testing is employed to mitigate the impact of erroneous results. These covariances are converted into the accuracy metric and utilized to construct the DLG to guide the localization enhancement. The place recognition-based RL method integrated with the DLG is shown as a case study. When the accuracy between all robots reaches 0.98, the APE is less than 0.1.

Regarding future work, on the one hand, we aim to integrate the DLG with multi-robot exploration of unknown environments to balance localization uncertainty and map uncertainty. On the other hand, it is also promising to further filter the candidate keyframes for localization enhancement according to the covariance of the global posterior distribution to reduce the number of relative localizations.

REFERENCES

- [1] C. Cao, H. Zhu, Z. Ren, H. Choset, and J. Zhang, "Representation granularity enables time-efficient autonomous exploration in large, complex worlds," *Sci. Robot.*, vol. 8, no. 80, p. eadfo970, 2023.
- [2] F. Pasqualetti, A. Franchi, and F. Bullo, "On Cooperative Patrolling: Optimal trajectories, complexity analysis, and approximation algorithms," *IEEE Trans. Robotics*, vol. 28, no. 3, pp. 592–606, 2012.
- [3] J. Liu and G. Hu, "Relative localization estimation for multiple robots via the rotating Ultra-Wideband tag," *IEEE Robot. Automat. Lett.*, vol. 8, no. 7, pp. 4187–4194, 2023.
- [4] P. Yin, S. Zhao, H. Lai, R. Ge, J. Zhang, H. Choset, and S. Scherer, "AutoMerge: A framework for map assembling and smoothing in city-scale environments," *IEEE Trans. Robotics*, vol. 39, no. 5, pp. 3686–3704, 2023.
- [5] Z. Ning, Y. Zhang, J. Li, Z. Chen, and S. Zhao, "A bearing-angle approach for unknown target motion analysis based on visual measurements," *Int. J. Robot. Res.*, vol. 43, no. 8, pp. 1228–1249, 2024.
- [6] S. Dong, K. Xu, Q. Zhou, A. Tagliasacchi, S. Xin, M. Nießner, and B. Chen, "Multi-robot collaborative dense scene reconstruction," *ACM Trans. Graph.*, vol. 38, jul 2019.
- [7] 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 Trans. Robotics*, vol. 38, no. 4, pp. 2022–2038, 2022.
- [8] H. Shen, Q. Zong, B. Tian, and H. Lu, "Voxel-based localization and mapping for multirobot system in gps-denied environments," *IEEE Trans. Ind. Electron.*, vol. 69, no. 10, pp. 10333–10342, 2022.
- [9] L. Bernreiter, S. Khattak, L. Ott, R. Siegwart, M. Hutter, and C. Cadena, "A framework for collaborative multi-robot mapping using spectral graph wavelets," *Int. J. Robot. Res.*, vol. 43, no. 13, pp. 2070–2088, 2024.
- [10] J. A. Placed, J. Strader, H. Carrillo, N. Atanasov, V. Indelman, L. Carlone, and J. A. Castellanos, "A survey on active simultaneous localization and mapping: State of the art and new frontiers," *IEEE Trans. Robotics*, vol. 39, no. 3, pp. 1686–1705, 2023.

- [11] 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 Robot. Automat. Lett.*, vol. 5, no. 2, pp. 1127–1134, 2020.
- [12] S. Garg, N. Suenderhauf, and M. Milford, "Semantic–geometric visual place recognition: a new perspective for reconciling opposing views," *Int. J. Robot. Res.*, vol. 41, no. 6, pp. 573–598, 2022.
- [13] P. Yin, F. Wang, A. Egorov, J. Hou, Z. Jia, and J. Han, "Fast sequence-matching enhanced viewpoint-invariant 3-d place recognition," *IEEE Trans. Ind. Electron.*, vol. 69, no. 2, pp. 2127–2135, 2021.
- [14] J. Kim, D.-S. Han, and B.-T. Zhang, "Robust map fusion with visual attention utilizing multi-agent rendezvous," in *Proc. IEEE Int. Conf. Robot. Autom.*, pp. 2062–2068, 2023.
- [15] P.-Y. Lajoie and G. Beltrame, "Swarm-SLAM: Sparse decentralized collaborative simultaneous localization and mapping framework for multi-robot systems," *IEEE Robot. Automat. Lett.*, vol. 9, no. 1, pp. 475–482, 2024.
- [16] S. Chen, Y. Li, and W. Dong, "High-performance relative localization based on key-node seeking considering aerial drags using range and odometry measurements," *IEEE Trans. Ind. Electron.*, vol. 71, no. 6, pp. 6021–6031, 2023.
- [17] G. Chen, K. Song, X. Xu, W. Liu, and Z. Xiong, "RHAML: Rendezvous-based hierarchical architecture for mutual localization," *IEEE Robot. Automat. Lett.*, vol. 9, no. 7, pp. 6440–6447, 2024.
- [18] J. A. Placed and J. A. Castellanos, "Fast autonomous robotic exploration using the underlying graph structure," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 6672–6679, 2021.
- [19] R. Bai, S. Yuan, H. Guo, P. Yin, W.-Y. Yau, and L. Xie, "Multi-robot active graph exploration with reduced pose-slam uncertainty via submodular optimization," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 10229–10236, 2024.
- [20] J. Yan, X. Lin, Z. Ren, S. Zhao, J. Yu, C. Cao, P. Yin, J. Zhang, and S. Scherer, "MUI-TARE: Cooperative multi-agent exploration with unknown initial position," *IEEE Robot. Automat. Lett.*, vol. 8, no. 7, pp. 4299–4306, 2023.
- [21] Z. Zhang, J. Yu, J. Tang, Y. Xu, and Y. Wang, "MR-TopoMap: Multi-robot exploration based on topological map in communication restricted environment," *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 10794–10801, 2022.
- [22] V. Indelman, "Cooperative multi-robot belief space planning for autonomous navigation in unknown environments," *Auton. Robots*, vol. 42, pp. 353–373, 2018.
- [23] F. Gu, Y. He, and J. Han, "Active persistent localization of a three-dimensional moving target under set-membership uncertainty description through cooperation of multiple mobile robots," *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 4958–4971, 2015.
- [24] M. J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *Int. J. Comp. Vis.*, vol. 19, no. 1, pp. 57–91, 1996.
- [25] A. Blake and A. Zisserman, *Visual reconstruction*. Cambridge, MA, USA: MIT Press, 1987.
- [26] F. Dellaert and G. Contributors, "Georgia tech smoothing and mapping (gtSAM)," 2019. [Online]. Available: <https://gtSAM.org/>.
- [27] J. G. Mangelson, M. Ghaffari, R. Vasudevan, and R. M. Eustice, "Characterizing the uncertainty of jointly distributed poses in the lie algebra," *IEEE Trans. Robotics*, vol. 36, no. 5, pp. 1371–1388, 2020.
- [28] R. Sim and N. Roy, "Global a-optimal robot exploration in slam," in *Proc. IEEE Int. Conf. Robot. Autom.*, pp. 661–666, 2005.
- [29] K. Song, W. Liu, G. Chen, X. Xu, and Z. Xiong, "FHT-Map: Feature-based hybrid topological map for relocalization and path planning," *IEEE Robot. Automat. Lett.*, vol. 9, no. 6, pp. 5401–5408, 2024.
- [30] F. Radenović, G. Tolias, and O. Chum, "Fine-Tuning CNN image retrieval with no human annotation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 41, no. 7, pp. 1655–1668, 2019.
- [31] F. A. Maken, F. Ramos, and L. Ott, "Stein icp for uncertainty estimation in point cloud matching," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 1063–1070, 2022.
- [32] Y. Li, Z. Li, N. Chen, M. Gong, Z. Lyu, Z. Wang, P. Jiang, and C. Feng, "Multiagent multitraversal multimodal self-driving: Open mars dataset," in *Proc. IEEE Conf. Comp. Vis. Pattern Recog.*, pp. 22041–22051, June 2024.
- [33] C. Zheng, W. Xu, Z. Zou, T. Hua, C. Yuan, D. He, B. Zhou, Z. Liu, J. Lin, F. Zhu, Y. Ren, R. Wang, F. Meng, and F. Zhang, "FAST-LIVO2: Fast, direct lidar-inertial-visual odometry," *IEEE Trans. Robotics*, vol. 41, pp. 326–346, 2025.



Gaoming Chen received the B.E. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2021. He is currently working toward the Ph.D. degree in mechanical engineering with the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China.

His research interests include multi-robot localization and multi-robot exploration.



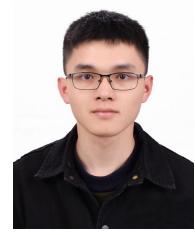
Kun Song received the B.E. degree in mechanical engineering from the the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2022. He is currently working toward the M.S. degree in mechanical engineering with the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China.

His research interests include multi-robot exploration and cooperative manipulation.

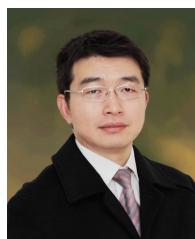


Wenhong Liu received the B.E. degree in mechanical engineering from the School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2020. He is currently working toward the Ph.D. degree in mechanical engineering with the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China.

His research interests include multi-robot motion planning and cooperative manipulation.



Wenyao Ma received the B.E. degree in mechanical engineering from the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China, in 2023. He is currently working toward the Ph.D. degree in mechanical engineering with the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include multi-robot manipulation and reinforcement learning.



Zhenhua Xiong (member, IEEE) received the B.E. and M.E. degrees in aircraft design and vibration engineering from the Department of Aircraft Design, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1995 and 1998, respectively, and the Ph.D. degree in mechatronics from the Department of Electrical and Electronics Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, in 2002.

He is currently a Professor with the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include multi-robot system and intelligent manufacturing.