Distributed Real-time Cooperative Localization and Mapping using an Uncertainty-Aware Expectation Maximization Approach

Jing Dong, Erik Nelson, Vadim Indelman, Nathan Michael, Frank Dellaert

***Abstract*— We demonstrate distributed, online, and real-time cooperative localization and mapping between multiple robots operating throughout an unknown environment using indirect measurements. We present a novel Expectation Maximization (EM) based approach to efficiently identify inlier multi-robot loop closures by incorporating robot pose uncertainty, which significantly improves the trajectory accuracy over long-term navigation. An EM and hypothesis based method is used to determine a common reference frame. We detail a 2D laser scan correspondence method to form robust correspondences between laser scans shared amongst robots. The implementation is experimentally validated using teams of aerial vehicles, and analyzed to determine its accuracy, computational efficiency, scalability to many robots, and robustness to varying environments. We demonstrate through multiple experiments that our method can efficiently build maps of large indoor and outdoor environments in a distributed, online, and real-time setting.**

I. Introduction

Distributed cooperative mapping in unknown environments is an important ability for a wide variety of multi-robot applications such as coordinated control, surveillance, and rapid exploration. Applications relying on distributed mapping must assume a shared environment representation in order to establish a common reference frame between vehicles. For this reason, two essential capabilities for a cooperative mapping system are the ability to robustly and efficiently compute data association, and establish relative pose transformations between individual vehicles in an online and distributed manner (Fig. 1).

Previous solutions to these problems can be divided into Full-SLAM and Pose-SLAM. In multi-robot Full-SLAM approaches, robots have shared views of common landmarks [1], [2], [3], [4], [5]. Several of these Full-SLAM methods have the ability to build robust multi-robot data associations without a prior on initial relative poses [2], [5]. Specifically, Fox et al. consider a similar problem which was solved using a particle filter approach in the Full-SLAM framework [6].

Alternatively, Pose-SLAM approaches achieve greater efficiency and robustness in comparison to Full-SLAM by avoiding explicit estimation of landmark positions. In this work we will focus on the Pose-SLAM approach. Many multi-robot Pose-SLAM techniques assume the existence of *direct* relative pose measurements between robots, and can compute a common reference frame with or without assuming prior information on initial relative poses between robots [7], [8]. However, these strategies make strong temporal and proximal assumptions regarding interactions between vehicles, which may not be viable for operation in expansive environments where vehicles navigate either asynchronously or without expectation of frequent rendezvous.

In contrast, we focus on missions through large and complex environments which do not permit frequent vehicle interaction, and assume that vehicles autonomously explore the environment while opportunistically coordinating to build a common reference frame. Because of these constraints, we require robots to localize to one another solely through *indirect* observations (i.e. through commonalities observed

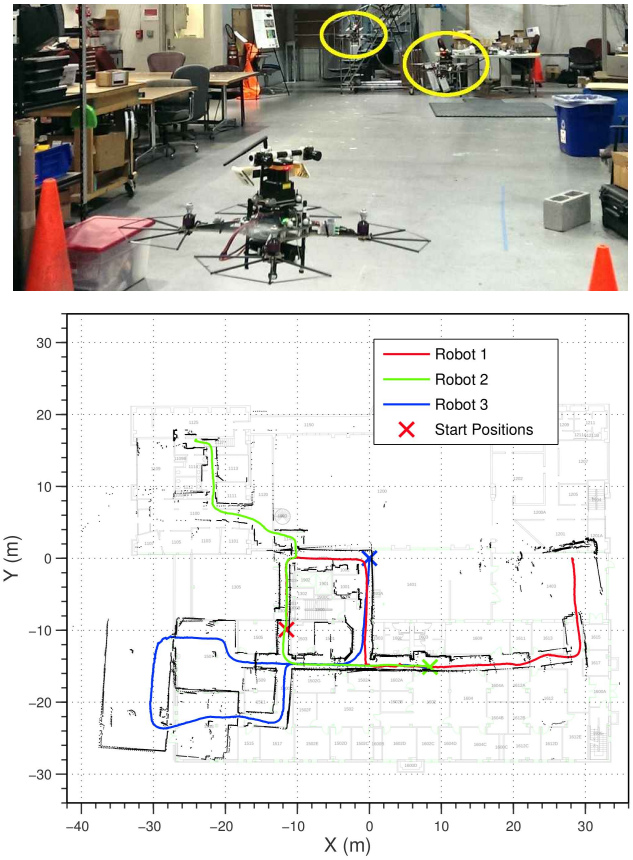


Fig. 1: Top: three quadrotors cooperatively generating a map of the environment. Bottom: Three robot trajectories and emplaced laser scan points in a computed common reference frame overlaid on a ground truth floor plan. Map misalignments caused by drift of odometries.

J. Dong and F. Dellaert are with the College of Computing, Georgia Institute of Technology, USA. {jdong, fd27}@gatech.edu E. Nelson and N. Michael are with the Robotics Institute, Carnegie Mellon University, USA. {enelson, nmichael}@cmu.edu V. Indelman is with the Aerospace Engineering, Technion - Israel Institute of Technology, Israel. vadim.indelman@technion.ac.il We gratefully acknowledge the support of ARL grants W911NF-08-2-0004 and W911NF-08-D-0004.

in collected sensory observations). The problem of building a common reference frame has been solved under the assumption of perfect multi-robot data association [9], [10], [11], [12], but these works do not allow for erroneous data association. To address this issue, methods based on Expectation Maximization (EM) [13], [14] which simultaneously compute a common reference frame and infer data association through *imperfect indirect* measurements have been proposed. Unfortunately, these approaches assume perfect trajectory estimates, and are unable to identify multi-robot data associations in the presence of drift.

In this paper, we significantly improve upon a previous EM-based Pose-SLAM approach [14] [15], focusing towards a *distributed*, *online*, and *real-time* implementation of multi-robot cooperative mapping with *unknown* initial poses and *indirect* data association. We detail an online EM optimization and hypothesis based method to estimate the multi-robot common reference frame. In addition, as our main contribution, we propose an EM-based online method to robustly identify multi-robot loop closures in pose graphs, which incorporates robot pose uncertainty to filter incorrect loop closures caused by trajectory drift. To comprehensively evaluate the system's accuracy, efficiency, scalability, and robustness of the approach, we detail online multi-robot mapping experiments conducted on aerial robots.

Implementation of our approach requires a fast 2D laser correspondence method to robustly detect correspondences between shared sensory observations. Multi-robot data association is similar in nature to finding loop closures in the single robot Pose-SLAM problem: both problems involve comparing a query scan to a set of cached scans. The distinction between the two is that the query scan in the former problem is received from a different robot. Therefore to compute sensory correspondences between robots, we borrow ideas from single robot 2D laser scan loop closure literature. Granstrom et al. [16] present a method which combines many global laser scan features, such as Centroid and Close Area features, to determine loop closures. However, this strategy is not view-point invariant which makes it unsuitable for a multi-robot localization task. Several works have developed laser-based feature detection and description methods [17], [18] that can be used to rapidly detect correspondences. We choose to use FLIRT features [18] introduced by Tipaldi et al. due to their rotational invariance.

The main contributions of this paper are:

- A novel EM-based technique for robust multi-robot loop closing which integrates robot pose uncertainty.
- A robust multi-robot laser scan correspondence strategy utilizing FLIRT features and RANSAC registration.
- A real-time implementation of our approach, and examination of its accuracy, efficiency, scalability, and robustness through experiments with aerial robots.

The approach can be explained through the following procedure: Robots are initialized and begin localizing and building individual maps. As they navigate, each robot shares locally acquired laser scans and SLAM pose estimates with other robots over a wireless network. FLIRT features are extracted from laser scans received by each robot and are compared against that robot's local history of laser scan features. RANSAC is used to find correspondences between the feature sets. We then use EM to cluster correspondences and detect inliers amongst those clusters. The set of inlier correspondence clusters is used to build multiple transformation hypotheses for the robot pair. The most probable hypothesis is used to initialize a pose graph transformation constraint between the robots, which is optimized over time using iSAM2 incremental optimizer [19].

II. Problem Formulation

We begin with a probabilistic formulation for a single robot r. Letting X^r represent the robot trajectory and denoting all observations obtained thus far by Z^r, the joint probability distribution function can be written as

$$p(X^r|Z^r) \propto p(x_0^r) \prod_{u_{k,l}^r \in Z^r} p(u_{k,l}^r|x_k^r, x_l^r), \tag{1}$$

with $p(x_0^r)$ being the prior on the initial pose x_0^r, and $x_i^r \in X^r$ representing the robot pose at time step i. We use the notation $u_{k,l}^r \in Z^r$ to represent a relative measurement involving two poses x_k^r and x_l^r, which could be produced, for example, by a camera or a range sensor.

Following the common assumption in the SLAM community, we define the single measurement likelihood term as a single Gaussian

$$p(u_{k,l}^r|x_k^r, x_l^r) \propto \exp\left(-\frac{1}{2}\left\|u_{k,l}^r \ominus (x_k^r \ominus x_l^r)\right\|_\Sigma^2\right), \tag{2}$$

where Σ is the measurement noise covariance and $x_k^r \ominus x_l^r$ expresses x_l^r in the frame of x_k^r. The single robot Pose-SLAM problem of robot r can be formulated as a MAP estimation of joint probability in Eq. (1)

$$\hat{X}^r = \underset{X^r}{\operatorname{argmax}}\, p(X^r|Z^r). \tag{3}$$

Extending this formulation to the multi-robot Pose-SLAM problem, we assume a group of R robots are deployed in an unknown environment to collaboratively explore and generate a shared map. This augments Eq. (3) to

$$\hat{\mathcal{X}}^{\mathcal{R}} = \underset{\mathcal{X}^{\mathcal{R}}}{\operatorname{argmax}}\, p(\mathcal{X}^{\mathcal{R}}|\mathcal{Z}^{\mathcal{R}}), \tag{4}$$

where $\mathcal{R} \doteq \{1, ..., R\}$ is the set of all robots, $\mathcal{X}^{\mathcal{R}} \doteq \{X^1, ..., X^R\}$ is the set of their trajectories, and $\mathcal{Z}^{\mathcal{R}}$ is the union of all single-robot measurements and all relative multi-robot measurements

$$\mathcal{Z}^{\mathcal{R}} \doteq \{Z^1, ..., Z^R\} \cup \{u_{k,l}^{r,r'} | (r, r', k, l) \in \mathcal{F}\} \tag{5}$$

Here, $\mathcal{F}$ denotes the set of multi-robot data associations, and $u_{k,l}^{r,r'}$ specifies a measurement correspondence between x_k^r and $x_l^{r'}$ for each data association $(r, r', k, l) \in \mathcal{F}$ between two robots r and r'. We note that the set of multi-robot data associations $\mathcal{F}$ may contain incorrect correspondences.

In contrast to the single robot scenario, since the trajectories of different robots are defined in their local frames,

the transformation between these reference frames must be considered. With this in mind, we define the likelihood term of the multi-robot constraint between robots r and r' as

$$p(u_{k,l}^{r,r'}|x_k^r, x_l^{r'}) \propto \exp\left(-\frac{1}{2}\left\|err(u_{k,l}^{r,r'}, x_k^r, x_l^{r'})\right\|_\Sigma^2\right) \quad (6)$$

with

$$err(u_{k,l}^{r,r'}, x_k^r, x_l^{r'}) = u_{k,l}^{r,r'} \ominus \left(x_k^r \ominus \left(T_{r'}^r \oplus x_l^{r'}\right)\right). \quad (7)$$

The notation $\oplus$ represents transform composition, and $T_{r'}^r$ is the transform between reference frames of robot r and r'.

Although the complete multi-robot Pose-SLAM problem is formulated in Eq. (4), solving this problem is not practical for real-time applications due to network and computational constraints limiting the frequency of shared sensor measurements. Instead, each robot r uses a subset of measurements and poses, $\mathcal{Z}^r \subseteq \mathcal{Z}^\mathcal{R}$ and $\mathcal{X}^r \subseteq \mathcal{X}^\mathcal{R}$, corresponding to all local measurements as well as a subset of measurements shared by other robots. Under this subsampling, the multi robot Pose-SLAM problem solved by robot r is

$$\hat{\mathcal{X}}^r = \underset{\mathcal{X}^r}{\operatorname{argmax}}\, p(\mathcal{X}^r|\mathcal{Z}^r) \quad (8)$$

In this paper we address the problem of recovering the MAP estimate in Eq. (8) while considering multi-robot data association and initial relative poses between the robots are both unknown. Similarly to [13], [14], we address the data association problem by introducing a binary variable $j_{k,l}^{r,r'} \in \mathcal{J}^r$ for each multi-robot relative pose measurement $u_{k,l}^{r,r'}$ that is available to robot r. These binary variables represents whether the measurement is an inlier or an outlier. To calculate the MAP inference in Eq. (8), we marginalize over these variables, resulting in:

$$\hat{\mathcal{X}}^r = \underset{\mathcal{X}^r}{\operatorname{argmax}} \sum_{\mathcal{J}^r} p(\mathcal{X}^r, \mathcal{J}^r|\mathcal{Z}^r). \quad (9)$$

III. EM-Based Multi Robot Inference

In this section we present our EM framework and discuss one of the main contributions of this paper. Instead of directly solving the inference problem in Eq. (9), our approach must first infer both the initial relative poses between the robots as well as multi-robot data association. As demonstrated in previous works [13], [14], these are two coupled problems that must be solved concurrently before attempting to solve Eq. (9), and can be formulated as

$$\hat{T_r^{r'}} = \underset{T_r^{r'}}{\operatorname{argmax}} \sum_{\mathcal{J}^r} p\left(T_r^{r'}, \mathcal{J}^r|\hat{\mathcal{X}}^r, \mathcal{Z}^r\right). \quad (10)$$

However, since the above inference involves accounting for all possible values for each of latent binary variables in $\mathcal{J}^r$, it is computationally intractable. We developed an efficient EM-based approach to perform this inference [13], [14].

In this paper we build upon prior research and contribute an important extension to identify multi-robot loop closures even when robot trajectories *significantly drift* over time, which was not addressed in [13], [14]. We first detail our previous approach (Sections III-A and III-B) and in Section III-C focus on the mentioned contribution.

A. Inference over Initial Relative Poses via EM

The initial relative pose between any two robots r and r' can be calculated from each candidate multi-robot correspondence $(r, r', k, l) \in \mathcal{F}^r$, involving the measurement $u_{k,l}^{r,r'}$, and poses x_k^r and $x_l^{r'}$ as follows:

$$T_r^{r'} = \left(x_k^r \oplus u_{k,l}^{r,r'}\right) \ominus x_l^{r'}, \quad (11)$$

where x_k^r and $x_l^{r'}$ are expressed in the local reference frame of each robot.

Calculating $T_r^{r'}$ for all correspondences in $\mathcal{F}^r$ and analyzing the distribution of the obtained transformations allows us to concurrently infer $T_r^{r'}$ and simultaneously establish data association (i.e. identify which correspondences in $\mathcal{F}^r$ are inliers and outliers). A key observation mentioned in [13] is that inlier correspondences will produce a cluster of similar transformations $T_r^{r'}$, while transformations calculated from outlier correspondences will be scattered. The remaining task therefore reduces to identifying the inlier cluster.

This concept is demonstrated in Fig. 2 (a) shows the estimated trajectories of two robots (green and red); these estimates were obtained using only local observations available to each robot, and the initial relative pose between the robots is unknown. The figure also shows the set $\mathcal{F}^r$ of candidate correspondences available to the green robot. The distribution of transformations $T_r^{r'}$, calculated via Eq. (11) for each such correspondence is shown in Fig. 2 (b). The large inlier cluster corresponds to the correct value of $T_r^{r'}$.

Identifying this cluster involves performing the inference in Eq. (10), which, as mentioned, is computationally intractable. The EM approach that we developed in [13] allows us to drastically reduce computational complexity while producing a locally-optimal solution.

We now summarize the EM formulation and elaborate on aspects that will be required later in Section III-C. To simplify notations, we remove the superscripts and subscripts r and r', and rewrite Eq. (10) as

$$\hat{T} = \underset{T}{\operatorname{argmax}} \sum_{\mathcal{J}} p\left(T, \mathcal{J}|\hat{\mathcal{X}}, \mathcal{Z}\right). \quad (12)$$

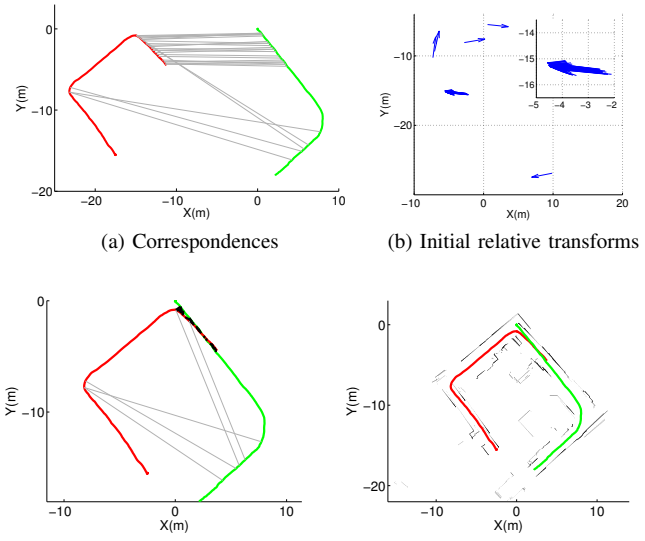
The tth iteration of the EM optimization can be described by the following 2 steps:

- **E step**: Calculate a lower bound on $p\left(T|\hat{\mathcal{X}}, \mathcal{Z}\right)$ via

$$\begin{aligned} Q\left(T|\hat{T}^{(t)}\right) &\doteq \mathbb{E}_{\mathcal{J}|\hat{T}^{(t)},\hat{\mathcal{X}},\mathcal{Z}}\left[\log p\left(T, \mathcal{J}|\hat{\mathcal{X}}, \mathcal{Z}\right)\right] \\ &= \sum_J p\left(\mathcal{J}|\hat{T}^{(t)}, \hat{\mathcal{X}}, \mathcal{Z}\right) \log p\left(T, \mathcal{J}|\hat{\mathcal{X}}, \mathcal{Z}\right). \end{aligned} \quad (13)$$

- **M step**: Calculate the MAP estimation by maximizing the above lower bound:

$$\hat{T}^{(t+1)} = \underset{T}{\operatorname{argmax}}\, Q\left(T|\hat{T}^{(t)}\right). \quad (14)$$



(a) Correspondences (b) Initial relative transforms

(c) Inliers and outliers (d) Trajectories after alignment

Fig. 2: The process of computing an initial relative transform. (a) shows correspondences between two trajectories. (b) depicts the transforms calculated through correspondences in (a) with an inlier cluster highlighted. (c) shows aligned trajectories with inliers and outliers in black and gray, respectively. (d) shows the resulting trajectories with scans.

Assuming the binary variables j_s in $\mathcal{J}$ are statistically independent conditioned on $\hat{T}^{(t)}, \hat{\mathcal{X}}$ and $\mathcal{Z}$, and letting $s \doteq (r, r', k, l) \in \mathcal{F}^r$ and $u_s \doteq u_{k,l}^{r,r'}$, Eq. (13) can be written as

$$Q\left(T|\hat{T}^{(t)}\right) = p\left(T|\hat{\mathcal{X}}^r\right) + \sum_{s\in\mathcal{F}^r}\sum_{j_s\in\mathcal{J}} p\left(j_s|u_s, \hat{T}^{(t)}, \hat{x}_k^r, \hat{x}_l^{r'}\right) \cdot \log p\left(j_s|\hat{x}_k^r, \hat{x}_l^{r'}\right) p\left(u_s|\hat{x}_k^r, \hat{x}_l^{r'}, j_s, T\right), \quad (15)$$

where

$$p\left(j_s|u_s, \hat{T}^{(t)}, \hat{x}_k^r, \hat{x}_l^{r'}\right) = \frac{p\left(u_s|\hat{x}_k^r, \hat{x}_l^{r'}, j_s, \hat{T}^{(t)}\right) p\left(j_s\right)}{p\left(u_s|\hat{T}^{(t)}, \hat{x}_k^r, \hat{x}_l^{r'}\right)}. \quad (16)$$

$p\left(j_s\right)$ is the prior term and the measurement likelihood $p\left(u_s|\hat{x}_k^r, \hat{x}_l^{r'}, j_s, \hat{T}^{(t)}\right)$ can be expanded as

$$p\left(u_s|\hat{x}_k^r, \hat{x}_l^{r'}, j_s, \hat{T}^{(t)}\right) \propto \exp\left(-\frac{1}{2}\left\|err(u_{k,l}^{r,r'}, \hat{x}_k^r, \hat{x}_l^{r'})\right\|_{\Sigma_{js}}^2\right), \quad (17)$$

The $err\left(.\right)$ function is defined in Eq. (7), and $\Sigma_{js} \in \{\Sigma_{inlier}, \Sigma_{outlier}\}$.

The denominator in Eq. (16) is not actually evaluated; instead, the numerator is calculated for both inlier and outlier, and the equation is normalized such that $p\left(j_s = inlier|u_s, \hat{T}^{(t)}, \hat{x}_k^r, \hat{x}_l^{r'}\right) + p\left(j_s = outier|u_s, \hat{T}^{(t)}, \hat{x}_k^r, \hat{x}_l^{r'}\right) = 1$.

In our implementation, we use a uniform prior $p(inlier) = p(outlier) = 0.5$, and set Σ_{inlier} to 0.2 m in position and 0.02 rad in rotation, respectively. $\Sigma_{outlier}$ is set to 500 m in position and 200 rad in rotation.

B. Hypothesis Selection

The nonlinear EM optimization in Eqs. (13)-(14) is guaranteed to converge to a *local* maxima of Eq. (12). Therefore, a reasonably good initial guess is required to recover the correct reference frame $T_r^{r'}$. It is for this reason that we choose several initial guesses and perform the EM optimization for each one of them. These initial guesses are obtained, in our implementation, by calculating a histogram and performing basic clustering on each element separately (x, y, and orientation), as further detailed in [13].

Performing EM optimization for each such initial guess produces a locally optimal solution, that we call a *hypothesis*, representing the inferred $T_r^{r'}$ and multi-robot data association. A hypothesis $h \doteq \{I, O\}$ is a partitioning of the set $\mathcal{J}^r$ into inliers I and outliers O, such that $I \cup O = \mathcal{J}^r$. The problem can be formulated now as choosing the most probable hypothesis $h^\star$ from the set $\mathcal{H}$. In this paper we reject any hypotheses in $\mathcal{H}$ with $p(j_s = inlier|u_s, \hat{T}^{(t)}, \hat{x}_k^r, \hat{x}_l^{r'}) < 0.8$ and then choose $h^\star$ according to the following criteria: (1) h^* has most inlier count compared to other hypotheses in $\mathcal{H}$, and (2) h^* has at least 10 inliers.

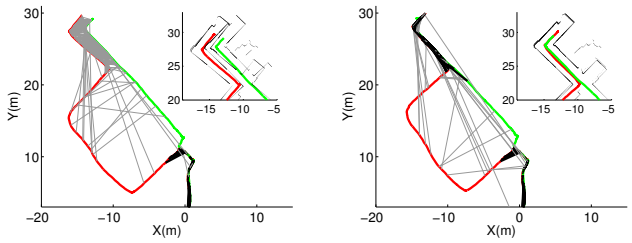
C. Identifying Multi-Robot Loop Closures Over Time

Once a common reference frame has been determined, it becomes possible for each robot r to perform the optimization Eq. (8) with its own measurements and the inlier multi-robot constraints. The latter is obtained by performing EM optimization and hypothesis selection (Sections III-A and III-B) when new multi-robot measurements are received.

However, due to drift in the robot pose estimates, identifying true multi-robot inlier correspondences as inliers becomes very challenging over time. The problematic aspect in the EM formulation from Section III-A is that it is conditioned on the robot estimated trajectories, without accounting for the uncertainties in these estimations, see Eq. (12). In practice, this is far from being correct, as the trajectory estimations develop significant drift over time.

An example is shown in Fig. 3(a), where the red and green robots start from same location, move towards the top-left, diverge, and finally reconverge in the same corridor. Upon reconvergence, due to significant drift in their trajectory estimations, the predicted relative pose will be far from the measured relative pose as determined by *true inlier* multi-robot correspondences. As a result, all of these inlier correspondences are identified as *outliers*, leading to sub-optimal map merging and estimation accuracy.

To address this issue, we incorporate the uncertainty in robot trajectories within the EM formulation from Section III-A, providing a probabilistically sound mechanism to adapt the inlier and outlier covariances $\Sigma_{inlier}, \Sigma_{outlier}$ from Eq. (17) according to the associated uncertainty in appropriate robot poses. In practice, $\Sigma_{outlier}$ needs not to be modified, as it is typically already very large.



(a) Without changing noise model (b) With changing noise model

Fig. 3: Two robots moving towards the top-left. (a) True relative pose measurements cannot be identified as inliers (black) due to poor trajectory alignment (inset). (b) With a changing noise model, inliers are correctly identified, correcting for drift.

For example, when considering the true inlier relative pose measurement $u_{k,l}^{r,r'}$ that relates between the poses x_k^r and $x_l^{r'}$, the uncertainties in estimates of these poses will be used to modify Σ_{inlier}. If these uncertainties are high (corresponding to a high estimation error) Σ_{inlier} will become large, and $u_{k,l}^{r,r'}$ will be identified as an inlier.

Specifically, the MAP estimation in Eq. (12) must now also be conditioned the joint covariance Σ:

$$\hat{T} = \underset{T}{\operatorname{argmax}} \sum_{\mathcal{J}} p\left(T, \mathcal{J} | \hat{\mathcal{X}}, \mathcal{Z}, \Sigma\right), \tag{18}$$

and the derivation in Eqs. (15)-(17) must be changed accordingly. Thus, the lower bound Eq. 15 becomes

$$Q\left(T|\hat{T}^{(t)}\right) = \sum_{J} p\left(\mathcal{J}|\hat{T}^{(t)}, \hat{\mathcal{X}}, \mathcal{Z}, \Sigma\right) \log p\left(T, \mathcal{J}|\hat{\mathcal{X}}, \mathcal{Z}, \Sigma\right), \tag{19}$$

and the explicit expression of this bound, given by Eq. (15), also becomes conditioned on Σ. In particular, the measurement likelihood changes from Eq. (17) to

$$p\left(u_s|\hat{x}_k^r, \hat{x}_l^{r'}, j_s, \hat{T}^{(t)}, \Sigma\right) \propto \exp\left(-\frac{1}{2}\left\|err(u_{k,l}^{r,r'}, \hat{x}_k^r, \hat{x}_l^{r'})\right\|_{\Sigma_v}^2\right), \tag{20}$$

which now includes a *modified* measurement covariance Σ_v.

We now focus on developing an expression for Σ_v. The predicted measurement u_s^{pred} can be written as

$$z_s^{pred} = \hat{x}_k^r \ominus \left(\hat{T}^{(t)} \oplus \hat{x}_l^{r'}\right) = h(\hat{T}^{(t)}, \hat{x}_k^r, \hat{x}_l^{r'}), \tag{21}$$

resulting in the following observation model:

$$u_s = h(T, \hat{x}_k^r, \hat{x}_l^{r'}) + v \tag{22}$$

with $v \sim N(0, \Sigma_v)$. Now, $T \equiv T_r^{r'}$ is the random variable, while the estimates $\hat{x}_k^r$ and $\hat{x}_l^{r'}$ are considered fixed (see Eq. (18)).

However, the inlier measurement noise covariance Σ_{inlier} only correctly models the error $u_s - h(T_r^{r'}, x_k^r, x_l^{r'})$ for *true* values $\bar{x}_k^r$ and $\bar{x}_l^{r'}$. Since x_k^r and $x_l^{r'}$ are fixed, any deviation from these values must be accounted for by the measurement covariance model Σ_v (see similar treatment in [12]).

We achieve this in the following way. To account for errors Δx_k^r and $\Delta x_l^{r'}$ in the estimates $\hat{x}_k^r$ and $\hat{x}_l^{r'}$, we write

$$\bar{x}_k^r = \hat{x}_k^r + \Delta x_k^r \ , \ \bar{x}_l^{r'} = \hat{x}_l^{r'} + \Delta x_l^{r'}. \tag{23}$$

Linearizing $h(T_r^{r'}, \bar{x}_k^r, \bar{x}_l^{r'})$ about $\hat{x}_k^r$ and $\hat{x}_l^{r'}$ yields:

$$u_s \approx h\left(T, \hat{x}_k^r, \hat{x}_l^{r'}\right) + \frac{\partial h}{\partial x_k^r}\Delta x_k^r + \frac{\partial h}{\partial x_l^{r'}}\Delta x_l^{r'} + v_{inlier}. \tag{24}$$

The uncertainty of the last three terms in Eq. (24) is quantified by Σ_v from Eq. (20), which is equal to:

$$\Sigma_v = \begin{pmatrix} \frac{\partial h}{\partial x_k^r} & \frac{\partial h}{\partial x_l^{r'}} \end{pmatrix} \Sigma_{x_k^r, x_l^{r'}} \begin{pmatrix} \frac{\partial h}{\partial x_k^r} \\ \frac{\partial h}{\partial x_l^{r'}} \end{pmatrix} + \Sigma_{inlier}, \tag{25}$$

where $\Sigma_{x_k^r, x_l^{r'}}$ is the joint marginal covariance of x_k^r and $x_l^{r'}$. This covariance can be extracted from the joint covariance Σ that represents the second moment of the pdf $p(\mathcal{X}^r|\mathcal{Z}^r)$ and can be evaluated using the MAP estimation Eq. (8).

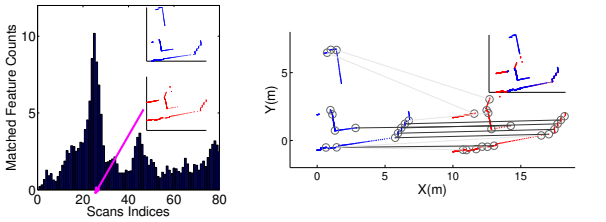
The approach described above also allows us to correctly determine multi-robot inlier correspondences even when the estimated trajectories develop significant drift. We use this approach in all of the experiments presented in this paper. For example, referring again to the scenario considered in Fig. 3, one can observe that this approach correctly identifies the inlier multi-robot correspondences, while the two robots travel in the same corridor (compare Figs. 3(a) and 3(b)), and as a result, the maps are well aligned.

IV. Feature-Based Correspondence

To implement the formulation in Section III, a practical multi-robot correspondence generation approach is required. In our previous works [13], [14], [15], naïve scan-to-scan ICP was used. However, ICP is both time consuming and susceptible to local minima in the presence of large rotations which are common when matching laser scans between robots in arbitrary reference frames. Instead we seek an efficient and robust method of forming correspondences. We propose a fast 2D laser scan based approach towards multi-robot data correspondence generation. The 2D laser correspondence generation problem can be formulated in the following manner: given a set of cached laser scans $\mathcal{L}^r \doteq \{L_i^r\}$ from robot r and a query scan $L_l^{r'}$ from robot r', determine the scans L_k^r in $\mathcal{L}^r$ which share similarity to $L_l^{r'}$, and align them to produce a relative pose measurement $u_{k,l}^{r,r'}$. Our approach can be divided into detection and matching. In the detection step, given a local scan set $\mathcal{L}^r$ and a received scan $L_l^{r'}$, we determine all correspondences (r, r', k, l). In the matching step, we align L_k^r with $L_l^{r'}$ and return the relative pose measurement $u_{k,l}^{r,r'}$.

A. Correspondence Detection

The detection approach can be divided into training and testing phases. To train, FLIRT [18] features and 128-dimensional descriptors are extracted for every incoming scan. The generated descriptors are inserted into randomized kd-trees provided by the FLANN library [20], granting the



(a) Feature histogram (b) RANSAC between matched features

Fig. 4: The loop closure process. (a) shows the feature match histogram as well as query and peak scans. (b) shows the matching phase, with features in circles, inlier and outlier feature matches in black and gray, respectively, and scans after alignment (inset).

ability to rapidly query for nearest neighbors in the 128-dimensional feature space.

In the testing phase, upon receiving a query scan from another robot, we first extract features and descriptors, then perform fast nearest neighbor searching in the local descriptor set for each query descriptor. Nearest neighbors of all query descriptors are placed in a histogram, indexed by local scan number. The peak value in the histogram corresponds to the local scan with the highest number of similar features, which implies a similarity between the two scans.

Figure 4 (a) shows an example query scan (blue) from another robot being matched against a set of 80 local scans. The scan at index 26 (red), corresponding to the histogram's peak, is indeed well-matched with the query scan.

B. Correspondence Matching

Once potential correspondences are identified, a RANSAC-based matching strategy is used to find the relative pose measurement for each loop correspondence. We first match FLIRT descriptors in each matched scan pair (query and local matched scan candidate) by finding the nearest neighbor, and use 2-point RANSAC to reject outliers and build a rigid relative transform. Figure 4 (b) depicts feature points plotted as circles, with nearest neighbor correspondences plotted as lines between features, and RANSAC inliers and outliers in black and gray, respectively. Finally the relative transform is refined by ICP.

V. Experimental Results

Teams of quadrotor were deployed into three indoor and outdoor environments to evaluate the accuracy, efficiency, scalability, and robustness of the approach. The quadrotor platforms were equipped with onboard computers (1.86 GHz Intel Core 2 Duo processor), IMUs, and 2D laser scanners. In addition to the incremental EM framework, each quadrotor estimated its trajectory using a laser and inertial filter-based 3D SLAM framework. Due to the fact that the individual 3D SLAM instances running on each quadrotor do not model loop closures, drift occurs over the trajectories.

We implement the multi-robot incremental EM algorithm using a distributed framework, where each robot shares laser scan and pose information over a wireless network with all other robots on the network. The GTSAM library[1] which includes an iSAM2 implementation is used for pose graph optimization. Our pipeline runs at a 10 Hz frequency on each robot, while laser scans are shared at lower frequency: a robot will share its most recent scan after moving 0.2 meters or after 1 second has transpired since the most recent scan share (triggered when either criteria is reached).

To evaluate the feasibility of our approach as well as its adaptability to varying environments, we provide an analysis of three online multi-robot trials. The first experimental environment is a large stretch of hallways with a central loop and a second loop through an outdoor patio. The second experiment consists of two robots initialized in separate buildings before switching buildings without observing one another directly. The third experiment takes place in a large outdoor hedge maze. Experiment 1 is shown in Fig. 1 and experiments 2 and 3 are depicted in Fig. 5.

A. Accuracy

The accuracy of the proposed approach was analyzed by comparing inter-robot transformations against ground truth measurements. In addition, robot trajectories in the computed common reference frame were validated through comparison against floor plans and satellite imagery.

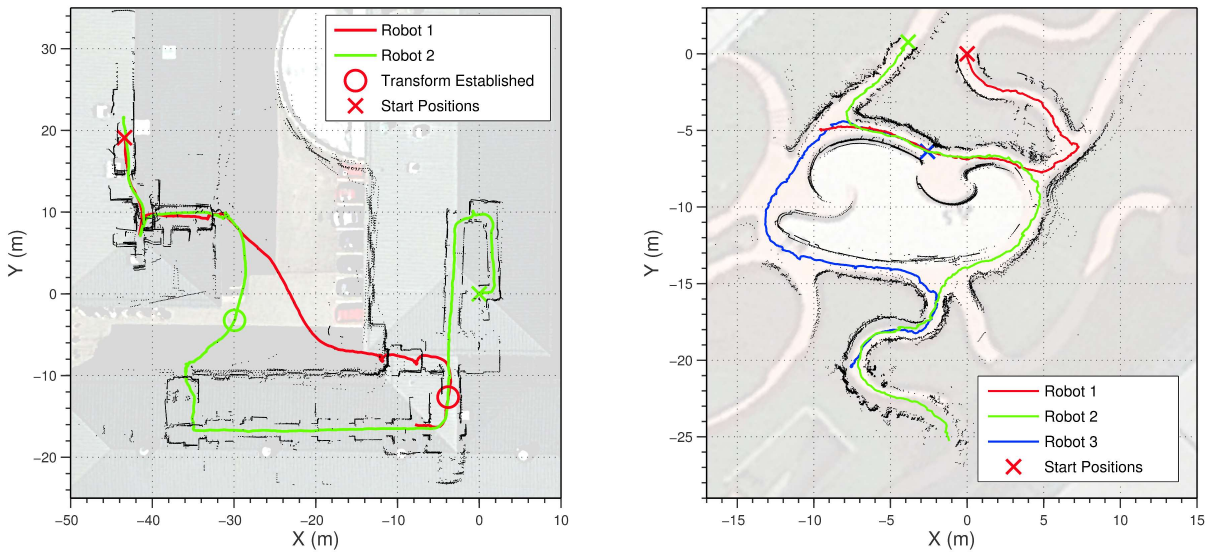
The initial estimated, final estimated, and ground truth between robot transformations estimated for experiment 1 are displayed in Table I, with ground truth ues measured by tape measure. The initial transformation estimate is built online during operation when a robot pair has highly weighted inlier correspondences, and is refined by online pose graph optimization to build the final transformation estimate at the end of the trial. Averaging across robots, the initial transformation error for experiment 1 is 0.82 m in translation and 0.07 rad in rotation. At the end of operation this error is reduced to 0.80 m in translation and 0.04 rad in rotation. The low initial error as well as the reduction in error with time implies that the EM-hypothesis based method is suitable for multi-robot mapping on this environmental scale.

Estimated trajectories and emplaced laser scans from the three experiments on top of a ground truth background are displayed in Fig. 1 and Fig. 5. While experiment 1 shows misalignment respect to the ground truth floor plan, this error manifests itself in areas with no sensory overlap and is therefore a consequence of drift in individual robot odometry. This drift can be bounded by further multi-robot observations. For instance, in Fig. 6, a single robot's trajectory was significantly improved through pose graph optimization after data associations were made between itself and other robots.

B. Efficiency

The CPU and usage of each robot was recorded during online experimentation to evaluate the efficiency of our algorithms in a real-time setting. The CPU usage of a robot from experiment 1 is shown in Fig. 7.

[1] https://collab.cc.gatech.edu/borg/gtsam/



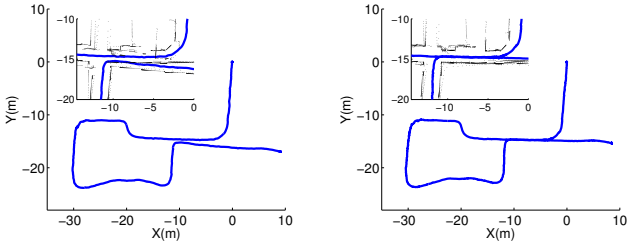
(a) Experiment 2: indoor and outdoor building exchange

(b) Experiment 3: outdoor hedge maze

Fig. 5: Aligned trajectories resulting from our approach for two outdoor experiments on top of satellite imagery. The points at which the two robots in experiment 2 established a common reference frame are marked with circles.

TABLE I: Initial relative transforms for experiment 1

		Robot 1		Robot 2		Robot 3	
		T_{r1}^{r2}	T_{r1}^{r3}	T_{r2}^{r1}	T_{r2}^{r3}	T_{r3}^{r1}	T_{r3}^{r2}
Original	x	-17.30	-0.26	11.29	-4.33	-2.49	16.27
	y	-11.04	-15.76	-17.13	-16.56	-15.55	-5.30
	θ	1.56	-3.00	-1.56	1.64	3.01	-1.64
Final	x	-17.46	-0.68	10.82	-4.36	-1.67	16.47
	y	-10.95	-15.19	-17.54	-16.75	-15.11	-5.32
	θ	1.58	-3.08	-1.58	1.63	3.08	-1.63
Measured	x	-16.80	-0.48	10.47	-5.32	-0.48	16.33
	y	-10.47	-15.80	-16.82	-16.33	-15.80	-5.32
	θ	1.57	-3.14	-1.57	1.57	3.14	-1.57



(a) Single robot SLAM trajectory. (b) After multi-robot EM optimization

Fig. 6: Accuracy improvement in the trajectory of a robot through multi-robot loop closures and pose graph optimization: robot 3, experiment 1. Map shown in inset image.

The overall CPU usage reaches a maximum at 62% during the experiment, and is generally below 30%. Our proposed approach consumes under 25% of the total CPU usage at all times, while other processes vary from 10% to 50% of CPU usage. When breaking our approach into individual algorithmic steps, FLIRT feature detection and description accounts for roughly 80% of our algorithm's computations. Prior to establishing relative transformations, hypothesis selection consumes a negligible amount of CPU. EM optimization has short bursts corresponding to inlier data associations between robots, which consume 10% of the total CPU at maximum. FLIRT feature extraction has roughly constant time complexity, so the main efficiency concern arises from expensive iSAM2 optimizations as the map grows, which is not the main focus of this work.

C. Scalability

Constrained network capacity is the most significant constraint limiting the scalability of our algorithm. In practice, packets containing laser scans are 11 kB in size, and are multicasted at an upper limit of 2 Hz. With pose estimates included the maximum transmission rate is 25 kB/s. Due to network latency, the maximum receive rate is 80 kB/s. Mesh networks exhibit $O(1/n)$ throughput when distributing packets between n systems [21]. Given a network with 11 Mb/s bandwidth and a maximum sharing rate of 2 Hz, this limits the number of active robots to 6-7. However, number of robots can be increased given a larger network bandwidth, lower latency and lower sharing rate.

D. Robustness

The three experiments detailed in Fig. 1 and Fig. 5 demonstrate that the approach is robust to structural symmetries in indoor environments (experiment 1), is unaffected by indoor to outdoor transitions as well as robots starting in different buildings (experiment 2), and is resilient to sparse and disparate laser scans in outdoor environments (experiment 3). The formulation assumes that robots communicate

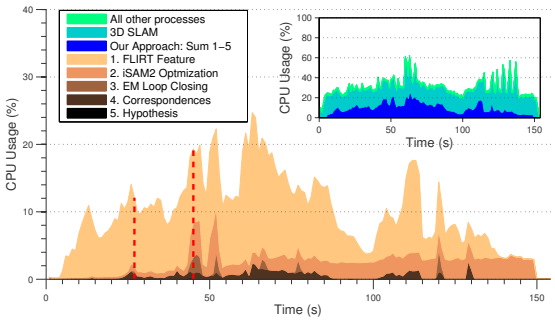


Fig. 7: CPU usage for robot 3, which had the longest deployment duration in experiment 1. Red lines mark the times when initial relative poses are selected.

opportunistically, and is therefore naturally insensitive to dropped packages and network latency.

VI. Discussion

Although the approach and its implementation have been comprehensively analyzed in Section V, several theoretical limitations to the formulation must be mentioned.

In Section III-B, hypotheses are chosen on the basis of number of inlier correspondences. While this criteria works in typical scenarios, including those presented in this work, it has potential to fail in the case of severe perceptual aliasing, i.e. situations where the environmental structure is exceedingly similar in two or more locations, which could result in choosing an incorrect hypothesis. A possible method of alleviating this issue is to require consideration of whether a sufficient amount of data has been accumulated to reliably perform hypothesis selection. We refer the reader to our previous work [14], which discusses these challenges.

Furthermore, our approach to incorporate uncertainty in robot trajectories within the EM formulation allows us to identify inlier multi-robot correspondences, even in the presence of significant trajectory drift. However, modifying the inlier covariance Σ_{inlier}, as described in Section III-C, could, theoretically, also result in identifying outlier correspondences as inliers if the error of these correspondences is in accordance with Σ_{inlier}. We note we have not observed this in the performed experiments. Nevertheless, addressing this aspect is the subject of future research.

VII. Conclusion

In this paper, we proposed a distributed and online approach to multi-robot cooperative localization and mapping. To address trajectory drift, we introduced a novel EM-based method that incorporates uncertainty to identify inlier and outlier multi-robot data associations. An EM-hypothesis based method was used to build a common reference frame. A robust feature-based multi-robot correspondence approach was presented. Our implementation was evaluated through online experiments in expansive environments, demonstrating that our solution is real-time viable, repeatable in varying environments, and scalable to larger fleets of robots.

References

[1] A. Howard, "Multi-robot simultaneous localization and mapping using particle filters," *Intl. J. of Robotics Research*, vol. 25, no. 12, pp. 1243–1256, 2006.

[2] L. Andersson and J. Nygards, "C-SAM : Multi-robot SLAM using square root information smoothing," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2008.

[3] K. Y. K. Leung, T. D. Barfoot, and H. H. T. Liu, "Decentralized cooperative slam for sparsely-communicating robot networks: A centralized-equivalent approach," *Journal of Intelligent & Robotic Systems*, vol. 66, no. 3, pp. 321–342, 2012.

[4] I. V. Melnyk, J. A. Hesch, and S. I. Roumeliotis, "Cooperative vision-aided inertial navigation using overlapping views," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 936–943, 2012.

[5] A. Cunningham, K. Wurm, W. Burgard, and F. Dellaert, "Fully distributed scalable smoothing and mapping with robust multi-robot data association," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, (St. Paul, MN), 2012.

[6] D. Fox, J. Ko, K. Konolige, B. Limketkai, D. Schulz, and B. Stewart, "Distributed multi-robot exploration and mapping," *Proceedings of the IEEE - Special Issue on Multi-robot Systems*, vol. 94, Jul 2006.

[7] X. Zhou and S. Roumeliotis, "Multi-robot SLAM with unknown initial correspondence: The robot rendezvous case," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pp. 1785–1792, IEEE, 2006.

[8] L. Carlone, M. K. Ng, J. Du, B. Bona, and M. Indri, "Rao-Blackwellized particle filters multi robot SLAM with unknown initial correspondences and limited communication," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 243–249, 2010.

[9] L. Merino, J. Wiklund, F. Caballero, A. Moe, J. M. De Dios, P.-E. Forssen, K. Nordberg, and A. Ollero, "Vision-based multi-uav position estimation," *Robotics & Automation Magazine*, vol. 13, no. 3, pp. 53–62, 2006.

[10] B. Kim, M. Kaess, L. Fletcher, J. Leonard, A. Bachrach, N. Roy, and S. Teller, "Multiple relative pose graphs for robust cooperative mapping," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, (Anchorage, Alaska), pp. 3185–3192, May 2010.

[11] V. Indelman, P. Gurfil, E. Rivlin, and H. Rotstein, "Graph-based distributed cooperative navigation for a general multi-robot measurement model," *Intl. J. of Robotics Research*, vol. 31, August 2012.

[12] V. Indelman, P. Gurfil, E. Rivlin, and H. Rotstein, "Distributed vision-aided cooperative localization and navigation based on three-view geometry," *Robotics and Autonomous Systems*, vol. 60, pp. 822–840, June 2012.

[13] V. Indelman, E. Nelson, N. Michael, and F. Dellaert, "Multi-robot pose graph localization and data association from unknown initial relative poses via expectation maximization," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2014.

[14] V. Indelman, N. Michael, and F. Dellaert, "Incremental distributed robust inference from arbitrary robot poses via em and model selection," in *RSS Workshop on Distributed Control and Estimation for Robotic Vehicle Networks*, 2014.

[15] E. Nelson, V. Indelman, N. Michael, and F. Dellaert, "An experimental study of robust distributed data association from arbitrary poses," in *Intl. Sym. on Experimental Robotics (ISER)*, 2014.

[16] K. Granstrom, J. Callmer, F. Ramos, and J. Nieto, "Learning to detect loop closure from range data," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 15–22, 2009.

[17] R. Zlot and M. Bosse, "Place recognition using keypoint similarities in 2d lidar maps," in *Experimental Robotics*, pp. 363–372, Springer, 2009.

[18] G. D. Tipaldi and K. O. Arras, "Flirt-interest regions for 2d range data," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pp. 3616–3622, 2010.

[19] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *Intl. J. of Robotics Research*, vol. 31, pp. 217–236, Feb 2012.

[20] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *International Conference on Computer Vision Theory and Application VISSAPP)*, pp. 331–340, 2009.

[21] J. Jun and M. L. Sichitiu, "The nominal capacity of wireless mesh networks," *Wireless Communications, IEEE*, vol. 10, no. 5, pp. 8–14, 2003.