Incremental Distributed Inference from Arbitrary Poses and Unknown Data Association

USING COLLABORATING ROBOTS TO ESTABLISH A COMMON REFERENCE

VADIM INDELMAN, ERIK NELSON, JING DONG, NATHAN MICHAEL, and FRANK DELLAERT

High-accuracy localization is a fundamental capability that is essential for autonomous reliable operation in numerous applications, including autonomous driving, monitoring of an environmental phenomena, mapping, and tracking. The problem can be formulated as inference over the robot's state and possibly additional variables of interest based on incoming sensor measurements and a priori information, if such information exists. Moreover, in numerous applications, this inference problem has to be solved in real time, thus requiring computationally efficient inference methods.

Localization approaches often consider the existence of absolute information from which the robot state can be determined. For example, numerous navigation algorithms rely on the Global Navigation Satellite System (GNSS) signal

Digital Object Identifier 10.1109/MCS.2015.2512031
Date of publication: 17 March 2016

1066-033X/16©2016IEEE

Overview of Graph-Based Simultaneous Localization and Mapping Algorithms

Simultaneous localization and mapping (SLAM) is a fundamental problem in robotics, consisting of robot (or sensor) localization while operating in environments that are partially or completely unknown, uncertain, or dynamically changing. As the robot moves, it observes the surrounding environment with its onboard sensors (for example, cameras or range sensors). Given these observations, the general concept is to concurrently estimate a map, representing the observed environment, while at the same time to also infer robot pose with respect to that map.

SLAM approaches can be classified into two types, pose SLAM and full SLAM, which differ in the variables to be inferred and in the way constraints are formulated. More specifically, the pose SLAM approach can be probabilistically formulated as

$$p_{\text{poseSLAM}}(X \mid Z) = p(x_0) \prod_{i,j} p(u_{i,j} \mid x_i, x_j), \tag{S1}$$

where $p(x_0)$ is the prior on the initial state x_0 and the term $p(u_{i,j} \mid x_i, x_j)$ represents a nonlinear constraint involving the robot states x_i and x_j, and a relative motion between these two states $u_{i,j}$. In the context of pose SLAM, the latter typically represents the relative motion calculated from raw observations z_i and z_j acquired at time instances t_i and t_j. For example, z_i and z_j can be laser scan observations, as considered in this article. It is important to note that the calculation of $u_{i,j}$ from z_i and z_j is done by an external process that is often treated as a black box (for example, ICP or image-based motion estimation).

The probabilistic formulation for full SLAM differs from (S1) and is

$$p_{\text{fullSLAM}}(X, L \mid Z) = p(x_0) \prod_i \left[p(u_{i,i-1} \mid x_i, x_{i-1}) \prod_{m \in \mathcal{M}_i} p(z_{i,m} \mid x_i, l_m) \right], \tag{S2}$$

where the set $\mathcal{M}_i$ represents landmark indices observed at time t_i, and for each such index m, $z_{i,m}$ is the actual observation (for example, image feature) and $l_m \in L$ is the corresponding landmark. In both cases, the formulation also includes a motion model $p(u_{i,i-1} \mid x_i, x_{i-1})$ that can be expressed in terms of the kinematics model of the robot and controls or can be represented instead by odometry, as is common in SLAM literature.

It is helpful to represent both inference problems using graphical models, such as a factor graph [S1]. Formally, a factor graph is a bipartite graph $G = (\mathcal{F}, \Theta, \mathcal{E})$ with two node types: variable nodes $\theta_j \in \Theta$ and factor nodes $f_i \in \mathcal{F}$. Edges $e_{ij} \in \mathcal{E}$ are always between factor and variable nodes.

Variable nodes in pose SLAM are robot poses X, while in full SLAM these nodes represent both robot poses X and landmarks L. Factor nodes represent the probabilistic terms in the corresponding formulations (S1) and (S2). It is therefore possible to rewrite the pose SLAM formulation in (S1) in terms of factors as

$$p_{\text{poseSLAM}}(X \mid Z) \propto \prod_i f_i(\Theta_i), \tag{S3}$$

and full SLAM formulation in (S2) as

$$p_{\text{fullSLAM}}(X, L \mid Z) \propto \prod_i f_i'(\Theta_i). \tag{S4}$$

The factor graphs represented by these two equations are illustrated in a simple example in Figure S1. Observe that since

that, together with additional sensors such as wheel odometry or inertial sensors, can be used in real time for calculating the navigation solution. Unfortunately, many scenarios exist where the GNSS is either unavailable or unreliable or where the obtained accuracy is insufficient. Indoor and urban navigation is one representative case for such scenarios.

Alternatively, localization approaches typically assume the map of the environment in which the robot has to operate is known a priori. Given such a map, it is possible to localize the robot by associating incoming sensor observations (for example from a laser range sensor or a camera) with the appropriate region in the map.

Yet, it is often the case that the robot has to operate in unknown, uncertain, or dynamically changing environments. In these scenarios, the map is thus either unavailable, only partially reliable, or has become partially obsolete as a result of a change in the environment. For example, numerous buildings collapsed in the Fukushima earthquake in 2011. The maps of these buildings, even if available, are of little use when sending robots for search and rescue missions and should be updated on the fly as the robots explore these environments.

Therefore, without reliable sources of absolute information (such as a GNSS signal or an a priori known valid map), solving the localization problem also requires mapping the environment in which the robot is to operate. This problem, known as simultaneous localization and mapping (SLAM), has been extensively investigated by the robotics research community in the past two decades. "Overview of Graph-Based Simultaneous Localization and Mapping Algorithms" provides an overview of the state-of-the-art approaches in SLAM.

Collaboration between multiple robots can yield higher levels of performance and is a key capability in multirobot autonomous systems that is of interest in a variety of problem domains, including surveillance, tracking, localization, mapping, and SLAM. Cooperatively inferring variables of interest, such as robot trajectories, observed objects, and tracked targets, results in higher levels of accuracy, flexibility, and robustness to failure. An illustrative example of multirobot inference is the experiment with autonomous flying vehicles shown in Figure 1. Robot trajectories and environment map are inferred in this example.

the underlying constraints in pose SLAM and full SLAM are different, different notations for factors are used. In the figure, the nodes x_i and l_j represent, respectively, robot state from time t_i and the jth landmark. Edges with different colors represent the different nonlinear constraints encoded by factors f_i (and f_i') in the original probabilistic formulations in (S1) and (S2).

A conceptual difference between the two formulations can be observed from Figure S1 and from (S1) and (S2). First, pose SLAM approaches only infer the robot past and current poses; on the other hand, full SLAM approaches also infer the landmarks L (representing the map). Second, full SLAM formulations typically use raw sensor observations (for example, image or range observations), while pose SLAM approaches first resort to black-box algorithms that calculate relative motion from the raw observations and then formulate corresponding constraints using the calculated motion. Because of the latter, pose SLAM does not correspond exactly to marginalization of landmarks in the general case

$$p_{\text{poseSLAM}}(X \mid Z) \neq p(X \mid Z) \equiv \int p_{\text{fullSLAM}}(X, L \mid Z)\, dL. \tag{S5}$$

Recently, accurate and computationally efficient approaches have been developed to solve the above inference problems. For example, to recover the MAP estimate for pose SLAM

$$X^* = \arg\max_X p_{\text{poseSLAM}}(X \mid Z). \tag{S6}$$

These approaches typically perform nonlinear iterative optimization while exploiting the sparsity of the involved matrices,

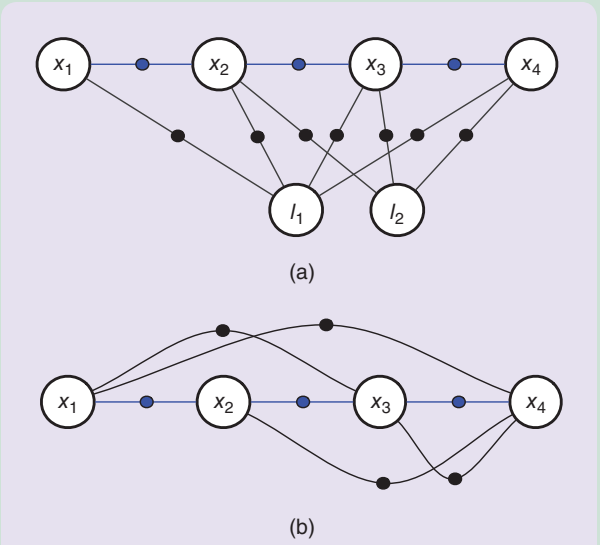


FIGURE S1 A factor graph representation of (a) full SLAM and (b) pose SLAM.

which can be naturally represented by a factor graph. For further details, the reader is referred to [S2] and [45].

REFERENCES

[S1] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[S2] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G^2o: A general framework for graph optimization," in *Proc. IEEE Int. Conf. Robotics Automation*, Shanghai, China, May 2011, pp. 3607–3613.

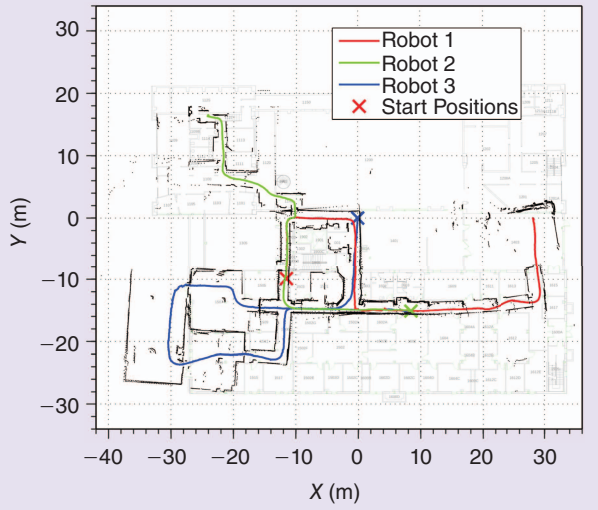
To facilitate cooperative inference, it is essential to establish a common reference frame and environment map between the robots so that robots can communicate relevant information with each other and correctly interpret that information for their needs. Only a few attempts have been made to simultaneously determine a common reference frame and resolve data association between sensor observations (for example, images or laser scans) acquired by different robots. Here, data association refers to determining whether different observations describe the same scene, where the scene is everything in the perceptual field of view of the robot (which varies based on the robot's sensors). For example, the scene could be the visual representation of the environment if the robot has a camera or the structure of the environment if the robot is equipped with a laser scanner.

Solving these two coupled problems is important because it enables the robots, which are scattered in a complex, initially unknown environment, to establish collaboration without requiring any prior knowledge or infrastructure. For example, starting from some arbitrary guess as to where each robot is and by sharing observations from onboard sensors, each robot can infer the trajectories of other robots in the group and, as a result, perform map merging.

Data association is a key challenge that shares some similarities with loop-closure detection in the single-robot case. Incorrect data association can lead to catastrophic deterioration in performance and must be avoided at all costs; the robotics community has been indeed very active in the last two years in developing robust graph-optimization techniques [1]–[5] to address this crucial aspect.

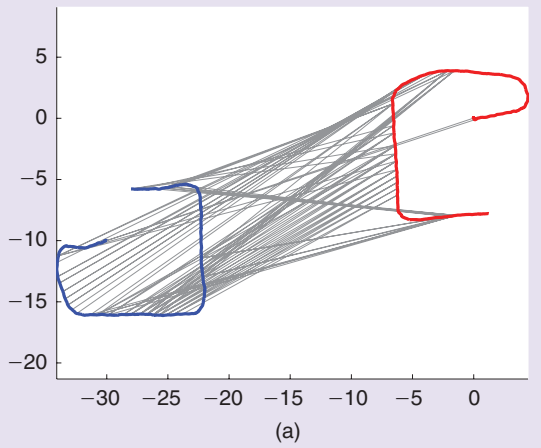
Multirobot data association has become an active research area as well [6]–[9], with the same sensitivity to incorrect correspondences as in the single-robot case. Multirobot data association, however, becomes more complicated when initial relative poses (positions and orientations) between the robots are unknown. An example is shown in Figure 2. Without a common reference frame, the robots must decide what information to share with each other. Calculated multirobot constraints based on this shared information can be used to distinguish false positives from correct correspondences between poses of multiple robots. Achieving these goals requires reasoning about multirobot data association and initial relative poses concurrently.

(a)

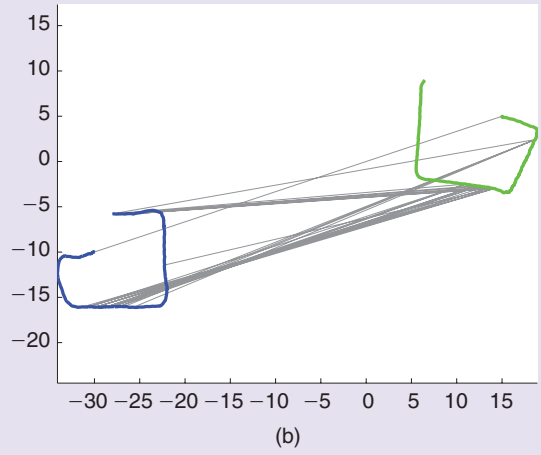


(b)

FIGURE 1 An example of multirobot inference. (a) Three quadrotors (yellow circles) cooperatively generating a map of the environment. (b) The output trajectories of three robots and the environment map of data set **D4**.



(a)



(b)

FIGURE 2 Multirobot candidate correspondences $\mathcal{F}^r$ between (a) red and blue and (b) green and blue robots in the real-world data set **D1**. Since a common reference frame is not yet established, robot initial poses are set to arbitrary values.

This article develops a distributed and incremental multirobot approach that allows a group of robots to concurrently establish a common reference frame and resolve multirobot data association on the fly. Furthermore, a probabilistic approach for estimating the positions of multiple robots is discussed.

The approach is based on the key observation that by analyzing the distribution of multirobot relative pose constraints [illustrated in Figure 3(b)], it is possible to estimate the transformation between the robot reference frames and identify the correct data association. Constraints on the poses between robots are formed by sharing and determining correspondences between sensor observations from different robots. The distribution of pose constraints generated in this fashion usually contains one *cluster* of correspondences for the correct transformation between robots and is scattered for outlier correspondences. However, sometimes *perceptual aliasing*, meaning cases in which two or more places are perceived as the same by the robot, can lead to multiple clusters corresponding to the correct transformation. For example, perceptual aliasing occurs when two identical corridors in the building lead to nearly identical sensor measurements. Perceptual aliasing introduces further complications, which are addressed by the approach presented in this article. Based on this insight, an expectation maximization (EM) [10] approach is developed to efficiently perform this inference by each of the robots independently, starting from several initial guesses and resulting in different locally optimal solutions.

Choosing the correct solution is a key challenge, since a wrong decision adds outliers to the graph optimization, and is particularly important when information is received *incrementally*, which is the case for the real-time operation in real autonomous robotic systems; in this case, a judgement must be made as to whether a *sufficient* amount of information has been received to perform this decision reliably.

Furthermore, perceptual aliasing presents additional challenges. Matching sensor observations from two different but

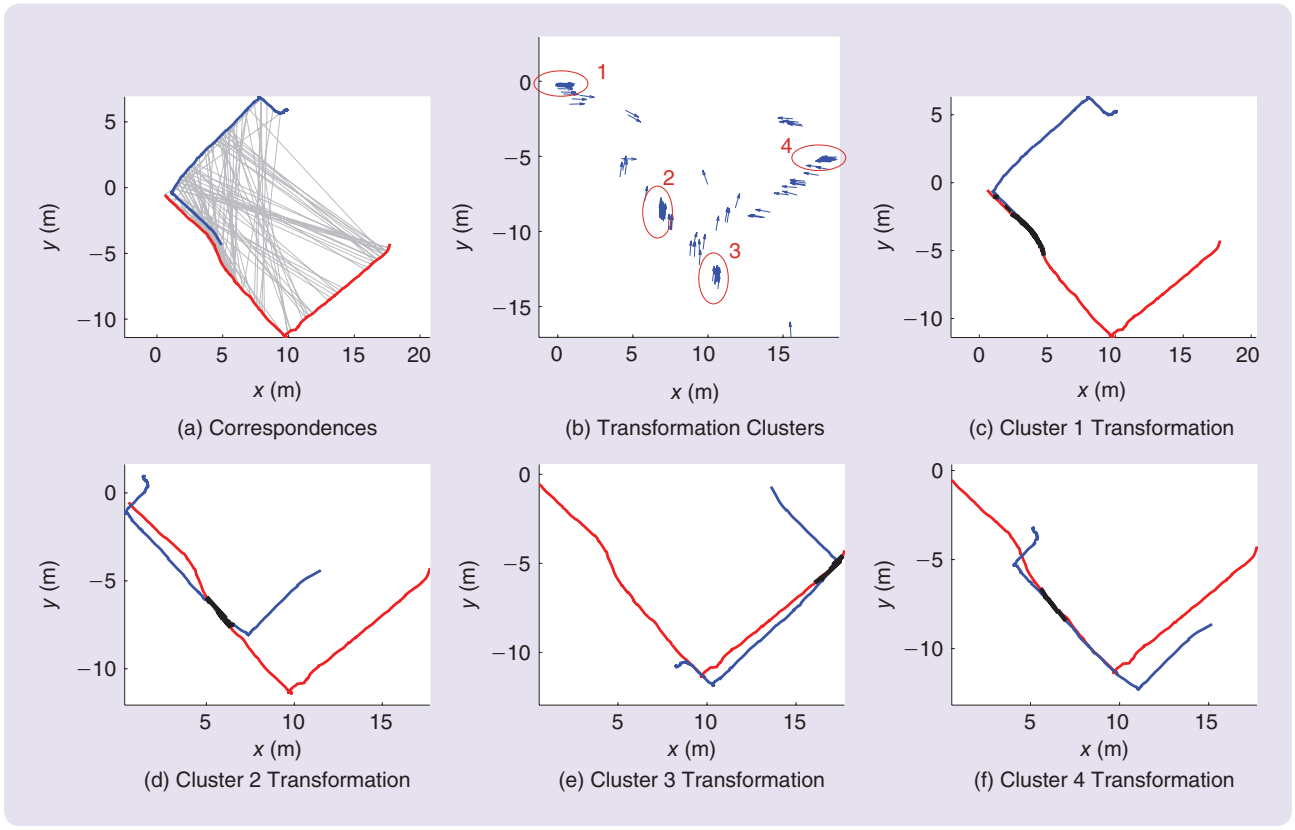


FIGURE 3 Measurement aliasing in data set **D3**. (a) shows the multirobot correspondences $\mathcal{F}^r$ found between the red robot and blue robot. Each correspondence is shown as gray line. (b) shows the calculated relative poses. Four clusters are marked by red circles and numbered from 1–4. One relative pose is generated per correspondence between two robots. Arrows indicate transformations from one robot to another, where each arrow's direction refers to a rotation and its origin refers to a translation. Cluster 1 represents the correct relative-pose transformation, and 2–4 are formed as a result of measurement aliasing. (c)–(f) show the aligned trajectories of red and blue robots of clusters 1–4, and the black lines show the correspondences in the cluster. (c) represents the correct transformation, and (d)–(f) are incorrect.

similar environments (such as two corridors) often results in reliable and consistent transformations that erroneously indicate the two environments are the same. These incorrect correspondences (referred to as *outliers*), and the correct correspondences (referred to as *inliers*) also form *clusters* of transformations between robots (see, for example, clusters 2–4 in Figure 3) that compete with the inliers cluster, leading to the problem of choosing the right cluster among several candidates.

This challenging problem is framed within a *model-selection* framework [11], to develop a probabilistically sound approach for selecting the most probable cluster. Moreover, the question of whether there is a correct cluster given the information available to each robot thus far is discussed, since the robots might not have observed the same environment yet. This problem is approached by modeling the prior probability for each cluster using the Chinese restaurant process (CRP) [12], [13]; see "Chinese Restaurant Process" for a brief overview.

Finally, this article considers challenges involved in operation in large-scale environments, where, due to drift accumulating in estimated robot poses, correctly identifying inlier multirobot constraints becomes even more difficult. To address this problem, a basic modification of the EM approach is described that performs well in practice in these challenging scenarios.

This article makes the following contributions: 1) the development of a new EM-based approach to determine initial relative poses and multirobot data association within an incremental distributed framework; 2) the development of a model-selection approach for identifying the correct solution among several candidates; in particular, this approach uses the CRP to cope with potential perceptual aliasing and to account for the possibility there is no correct solution given information obtained thus far; and 3) extensive evaluation in real-world multirobot experiments both in indoor and outdoor environments. In addition, an extension of the EM-based approach to large-scale scenarios, where robot trajectories develop significant drift, is suggested.

This article is an extension of [9] and [14]–[16]. An EM-based approach to solve the multirobot SLAM problem was first developed in [9], considering a centralized batch framework and without accounting for perceptual aliasing and the additional complexities of an incremental distributed

Chinese Restaurant Process

According to the Chinese restaurant process (CRP), the probability of observing a new place, after previously observing n unique places, is

$$\frac{\alpha}{\alpha + n - 1}, \tag{S7}$$

where α is a concentration parameter. This parameter defines the extent to which repeated observations of the same place take place; a larger α corresponds to higher probability of observing a *new* place. See Figure S2(a), which describes the evolution of the probability in (S7) as a function of n and α. The probability decreases with n for a given α.

Based on (S7), it is straightforward to show [43] that, given n observations, the probability of observing n unique places can be written as

$$f(n) \doteq \frac{\alpha^n}{\prod_{j=1}^{n} (\alpha + j - 1)}. \tag{S8}$$

Figure S2(b) illustrates this probability as a function of n and for different values of α. This probability decreases with n for a given α, corresponding to the above assumption that observing only new places becomes less likely as the robot makes more observations.

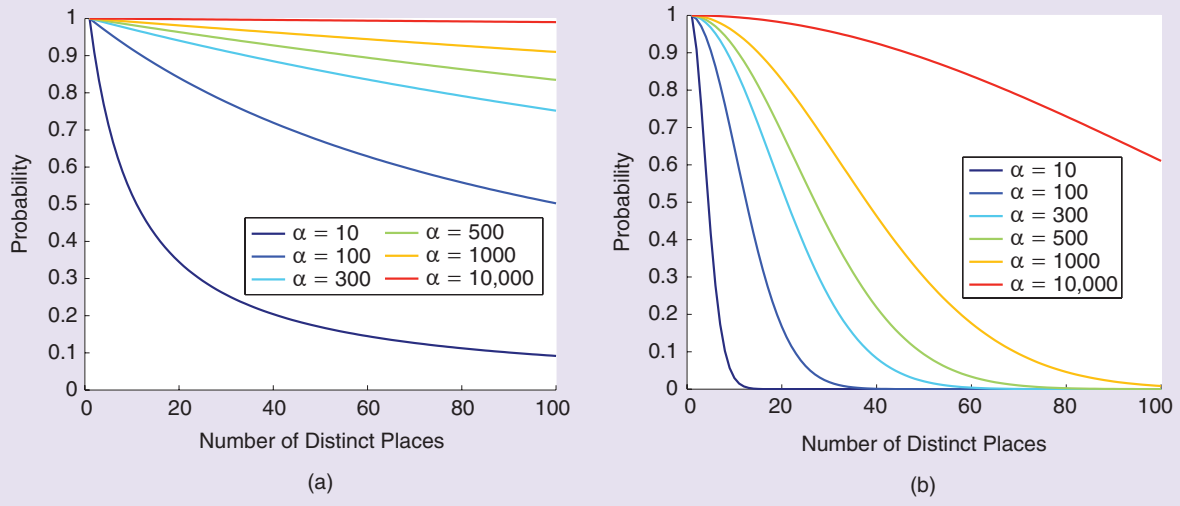


FIGURE S2 (a) The probability of observing a new place given n observations (b) the probability of observing n unique places given n observations.

problem setup. The approach was then extended to incremental and distributed configuration with real-time implementation [14], [16]. Several experimental techniques such as subsampling using autocovariance were also investigated in [15] to improve computational efficiency. A preliminary version of the current article appeared also in [14], which outlined the approach to cope with perceptual aliasing in a distributed incremental framework considered herein.

This article proceeds as follows. First, related work is discussed, and a high-level overview of the developed approach is provided. Next, the considered problem is formulated, and the EM-based framework to simultaneously infer initial relative poses and multirobot data association is introduced. The following section then focuses on hypothesis-model-based selection, which is an essential component to facilitate incremental operation in (potentially) perceptually aliased unknown environments. An analysis of computational complexity is then provided, followed by an extensive performance evaluation. Finally, approach limitations are discussed and suggestions for possible future research directions are provided.

RELATED WORK

Distributed cooperative localization and multirobot SLAM has been extensively studied over the last decade. Many of these research efforts, which can be classified into full SLAM [17]–[20] and pose SLAM [18], [21]–[25] approaches, assume that the initial relative poses between the robots are known and multirobot data association has been externally established. "Overview of Graph-Based Simultaneous Localization and Mapping Algorithms" provides an overview on graph-based SLAM algorithms and discusses the differences between pose and full SLAM formulations.

Several approaches have been developed to operate when the initial relative poses between the robots are unknown, assuming that perfect multirobot data association is given. These approaches, including [26]–[28], often use direct relative-pose measurements between the robots and

assume the identity of the observed robots is known. Relaxing this assumption has been investigated in [29]. Multirobot pose SLAM using *indirect* relative-pose constraints, corresponding to different robots observing the same environment, has been mainly researched under the assumption of perfect data association and an established common reference frame between the robots [30], [21], [23], [24].

Determining data association is often treated as a preprocessing step separate from inference over the robot states. In [6], an approach for distributed data association was developed based on random sample consensus (RANSAC) [31], and further a distributed framework computing global correspondences by identifying inconsistent data association is developed [7]. While these approaches do not require a common reference frame between the robots, they operate within the full SLAM paradigm, explicitly performing inference also over the observed structure (for example, three-dimensional points or objects) in addition to robot states.

In contrast, the following approach is formulated in the pose SLAM framework, which has computational advantages over full SLAM. Consistency-based outlier rejection has also been developed for multirobot pose SLAM [8] assuming direct relative-pose measurements, still assuming a common reference frame between the robots is known.

Another related line of research is multirobot map merging [32]–[34], where the aim is to integrate maps produced by different robots, typically without assuming a known common reference frame between the robots. For example, a topological approach that merges maps is developed in [34] by using a probabilistic generalized Voronoi diagram representation that accounts for the confidence levels (uncertainties) of different areas in the local maps. However, map-merging approaches do not typically reason about potential outliers in multirobot data association and have not been investigated, to the best of the authors' knowledge, in an incremental framework.

To be resilient to outliers overlooked by data-association approaches, the robotics community has been focusing on robust graph-optimization techniques [1]–[5], including the application of EM [10] for the single-robot case [4]. These new approaches, in particular, aim to be robust to loop-closure outliers since current state-of-the-art methods for generating loop-closure constraints (for example, FAB-MAP [35] place recognition) are not error free. However, the majority of robust graph-optimization approaches are developed for the single-robot case. In contrast, this article focuses on a multirobot distributed and incremental problem formulation. The incremental component introduces additional challenges, such as determining whether sufficient data has been accumulated for resolving multirobot data association, in particular, in the presence of perceptual aliasing. It is also noted that since, initially, the robots do not have a common reference frame established, the data-association problem is in particular challenging.

TABLE 1 The main notations used throughout the article.

Symbol	Description
R	Number of robots
$\mathcal{R}$	Set of all robots: $\mathcal{R} \doteq \{1,\ldots,R\}$
x_i^r	Pose of robot r at time t_i
X^r	Trajectory (in terms of poses) of robot r: $\{x_i^r\}$
$\mathcal{X}^{\mathcal{R}}$	Trajectories of all robots in the group: $\cup_{r=1}^{R} X^r$
$\mathcal{X}^r$	Variables inferred by robot r: X^r and additional variables of interest (for example, trajectories of other robots)
$u_{l,l'}^{r,r'}$	Multirobot constraint calculated from raw observations z_l^r and $z_{l'}^{r'}$
z_i^r	Raw sensor observation (for example, a laser scan) obtained by robot r at time t_i
Z^r	Local observations of robot r
$\mathcal{Z}^{\mathcal{R}}$	Observations of all the robots in the group
$\mathcal{Z}^r$	Observations available to robot r: Local observations Z^r and observations shared by other robots
$\mathcal{F}^{\mathcal{R}}$	Set of multirobot correspondences $\{(r,r',l,l')\}$
$\mathcal{F}^r$	Set of multirobot correspondences $\{(r,r',l,l')\}$ available to robot r
$j_{l,l'}^{r,r'}$	Binary variable, models whether the correspondence (r,r',l,l') is inlier or outlier
$\mathcal{J}^r$	Set of all binary variables $j_{l,l'}^{r,r'}$ that correspond to correspondences $\mathcal{F}^r$
$T_r^{r'}$	Transformation between the reference frames of robots r and r'
h	Hypothesis, defined in terms of inliers I and outliers O such that $I \cup O = \mathcal{J}^r$
$\mathcal{H}$	Set of all hypotheses

APPROACH OVERVIEW

The approach considers a group of R robots deployed to collaboratively operate in an unknown environment, initially unaware of each other. The objective is for each robot r to estimate its own trajectory X^r (current and past poses) and additional variables of interest, such as the trajectories of other robots, in a distributed incremental framework. Such a capability is important for multirobot cooperation in numerous applications; additionally, it allows the robots to extend their sensing horizon and establish a common environment map, observed so far by the entire group. Although the common map is not explicitly inferred, it can be always recovered from the estimated poses and sensor observations [36].

The main notational conventions used in this article are summarized in Table 1. Superscripts and subscripts denote, respectively, a robot's identity and time indices. For example, x_i^r represents the pose of robot r at time t_i.

In a centralized problem setting, this problem corresponds to calculating the maximum a posteriori (MAP) estimate

$$\hat{\mathcal{X}}^{\mathcal{R}} = \arg\max_{\mathcal{X}^{\mathcal{R}}} p(\mathcal{X}^{\mathcal{R}} | \mathcal{Z}^{\mathcal{R}}), \tag{1}$$

where $\mathcal{R} \doteq \{1, \ldots, R\}$, and $\mathcal{X}^{\mathcal{R}} \doteq \cup_{r=1}^{R} X^r$ and $\mathcal{Z}^{\mathcal{R}}$ represent the poses and the sensor observations of all the robots in the group. Here, X^r represents the trajectory of robot r in terms of past and current poses.

The robots are assumed to share sensor observations of different parts of the environment with each other, which, given data association, can be used to formulate constraints between appropriate poses of different robots. These constraints become part of the underlying graph structure of $p(\mathcal{X}^{\mathcal{R}} | \mathcal{Z}^{\mathcal{R}})$ and are essential for facilitating collaboration between the robots.

Each such constraint $u_{l,l'}^{r,r'}$ can involve two poses x_l^r and $x_{l'}^{r'}$ of two different robots r and r' at (potentially) different time instances represented by the indices l and l'. Denoting by $\mathcal{F}^{\mathcal{R}}$ the multirobot correspondences (r, r', l, l') between all robots in the group, and letting $\mathcal{Z}^r$ represent the local observations of each robot r, $\mathcal{Z}^{\mathcal{R}}$ in (1) is

$$\mathcal{Z}^{\mathcal{R}} \doteq \cup_{r=1}^{R} Z^r \cup \{u_{l,l'}^{r,r'} | (r, r', l, l') \in \mathcal{F}^r\}. \tag{2}$$

Assuming a common reference frame is known, the robots can identify and share with each other only observations of areas that are likely to be observed by other robots. However, without a common reference frame, it is not obvious what information the robots must communicate with each other since each robot r represents its trajectory X^r in its own local frame. Consequently, data association becomes much more challenging because of the high number of outlier correspondences, and the problem becomes is coupled with establishing a common reference frame between the robots.

A distributed setting further complicates this problem, since each robot r has only access to $\mathcal{Z}^r \subseteq \mathcal{Z}^{\mathcal{R}}$, composed of its own observations and observations shared by other robots. Letting $\mathcal{X}^r \subseteq \mathcal{X}^{\mathcal{R}}$ represent the trajectory of robot r and additional variables of interest, in this case, the appropriate trajectories of other robots (as defined below), the inference solved by each robot r is

$$\hat{\mathcal{X}}^r = \arg\max_{\mathcal{X}^r} p(\mathcal{X}^r | \mathcal{Z}^r). \tag{3}$$

An EM-based approach for solving this problem in an incremental and distributed setting with unknown multirobot data association and initial relative poses is described below. Since this approach requires the initial relative poses to be first roughly determined, the first part of the approach allows each robot to infer the initial relative poses and multirobot data association simultaneously. Directly solving the corresponding inference problem is computationally intractable because it involves introducing binary latent variables to model whether each individual multirobot correspondence is inlier or outlier and jointly accounting for all the possible values each of these variables can assume. In contrast, the developed EM approach enables significant reduction in computational complexity. The information available to each robot in an incremental problem setting at a given time often supports *several* possible solutions. The problem then turns into choosing the most probable solution and identifying whether a sufficient amount of information has been accumulated to make this choice reliably. The problem of choosing the most probable solution is particularly important in the incremental setting in the presence of *measurement aliasing*. A model selection-based paradigm is used to address this issue.

After the relative transformation between robots is computed, robot trajectories are inferred while continually using the incoming multirobot correspondences. Inlier correspondences are incrementally accepted using an EM approach. An extension is suggested to cope with significant trajectory drift by introducing modeling of pose uncertainty within EM.

Figure 4 shows a diagram of the approach. The input to the algorithm is the set of sensor observations from the local and remote sensors. Remote sensor observations are obtained by filtering and sharing observations subject to saliency and geometric criteria, such as the ability to rigidly register the sensor measurement with itself under a perturbation (one measure of saliency) or distance between captures. Local and remote sensor observations are compared and used to generate multirobot correspondences. Depending on whether the relative transformation is built or not, the workflow is switched: If the relative transformation is not determined, an EM optimization is performed from different initial guesses (for example, in Figure 3, four clusters represent four initial guesses). Each EM optimization converges to a local minimum, and each minimum represents a single *hypothesis* [in Figure 3, parts (c)–(f) represent four hypotheses, in which (c) is the correct one]. A model selection framework is used to determine the correct hypothesis once enough information is available. If the relative transformation has been determined already, EM optimization is performed using input correspondences. Identified inliers are inserted into a pose graph that is optimized. The output of the algorithm is the set of optimized poses and the identification of inlier and outlier correspondences.

PROBABILISTIC FORMULATION

A Bayesian formulation is developed here for the inference problem (3) to be solved by each robot r. The formulation begins with a probabilistic formulation for a single robot r and is extended to a multirobot incremental distributed framework.

Single-Robot Formulation

Denoting by Z^r the available sensor observations from robot r's onboard sensors, the joint probability distribution function (pdf) can be written as

$$p(X^r|Z^r) \propto p(x_0^r) \prod_{u_{k,l}^r \in Z^r} p(u_{k,l}^r|x_k^r, x_l^r), \tag{4}$$

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, both expressed relative to some local reference frame.

The measurement likelihood term $p(u_{k,l}^r|x_k^r, x_l^r)$ in (4) involves the relative-pose measurement $u_{k,l}^r \in Z^r$. This measurement can be directly calculated from vision or laser sensor observations at the two time instances t_k and t_l or can be obtained from odometry measurements. This measurement likelihood term is therefore used to represent both sequential pair-wise and loop-closure constraints, depending on the involved time instances. Following a standard assumption in SLAM community, this term is modeled as a Gaussian

$$p(u_{i-1,i}^r|x_{i-1}^r, x_i^r) \propto \exp\left(-\frac{1}{2}\|u_{i-1,i}^r \ominus h(x_{i-1}^r, x_i^r)\|_\Sigma^2\right), \tag{5}$$

where Σ is the measurement noise covariance, and h is a measurement model. In the case where robot poses are expressed in the same reference frame, this measurement model is $h(x_{i-1}^r, x_i^r) \doteq x_{i-1}^r \ominus x_i^r$. The notation $\ominus$ in $a \ominus b$ is used to express b locally in the frame of a. The MAP estimate of the rth robot poses X^r using only local information is

$$\hat{X}^r = \arg\max_{X^r} p(X^r|Z^r). \tag{6}$$

Multirobot Incremental Distributed Formulation

In a distributed incremental problem setting, the inference problem to be solved by each robot r is (3). Before proceeding with this inference problem, however, an approach for sharing information between robots is discussed for the case when a common reference frame between the robots is unknown.

Considering a distributed setting, the set of multirobot data association available to robot r is denoted $\mathcal{F}^r$, with each individual data association $(r, r', l, l') \in \mathcal{F}^r$ representing a relative-pose constraint $u_{l,l'}^{r,r'}$ relating between the pose of robot r at time t_l and the pose of robot r' at time $t_{l'}$. This constraint can represent both the direct measurement of one robot pose relative to another robot and also the estimated relative pose based on sensor observations of a common scene by two robots. In the latter case, it is computed from the sensor observations of the two robots $z_l^r \in Z^r$ and $z_{l'}^{r'} \in Z^{r'}$, which can represent, for example, laser scans or image observations.

Since a common reference frame between the robots is unknown, an approach is required for information sharing. Each robot r' is assumed to share at each time $t_{l'}$ its current sensor observation $z_{l'}^{r'}$, if it is informative, and also

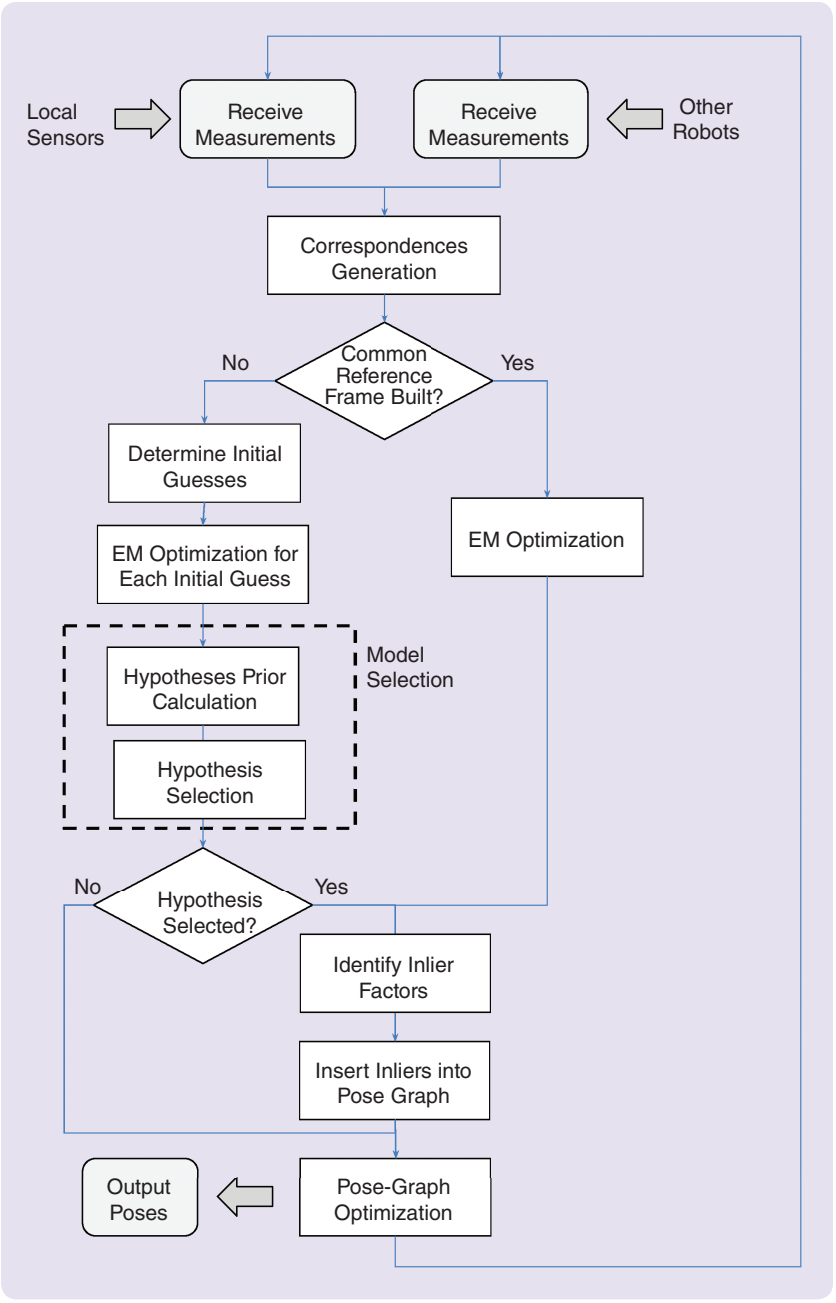


FIGURE 4 A diagram of the approach.

tracks all these informative observations $\{z_i^{r'}\}$ over time. Different methods can be used for quantifying the informativeness of a sensor observation. As discussed in detail later, in this article the observation saliency is used as a measure of informativeness; only highly salient observations are shared with other robots. Moreover, the observation saliency calculation and actual information sharing is done at a fixed frequency or upon moving a sufficient distance (0.5 m in the current implementation).

Additionally, robot r' shares local observations from its onboard sensors, which correspond to individual terms of $p(X^{r'}|Z^{r'})$ in (4). Using that information it is possible for robot r to recover the poses of robot r' using (6). While the amount of shared information may seem large, in practice these measurements can be represented by a single relative-pose constraint with an appropriate covariance [37]. Of course, if the robots only briefly interact, they might be capable of sharing only parts of their trajectories, instead of $p(X^{r'}|Z^{r'})$. In particular, this strategy would involve only sharing the probability distributions over robot poses directly involved in the multirobot measurements, while marginalizing out the rest of the variables. See [37] for further details and note that, while [37] considers a multirobot full SLAM framework, a similar approach can be used for the pose SLAM formulation in this article. A related approach was also described in [23] in the context of collaborative vision-aided navigation. While sharing only part of the variables is supported in implementation, for simplicity of the derivations to follow, the entire joint pdf $p(X^{r'}|Z^{r'})$ is assumed to be shared by each robot r'. Following this simplification, the random variables $\mathcal{X}^r$ that are inferred by each robot r are

$$\mathcal{X}^r \doteq \bigcup_r X^r. \tag{7}$$

Communicating only local observations, or the corresponding marginal $p(X^r|Z^r)$, guarantees *consistent* estimation because double counting is avoided [38], [37]. In other words, each raw sensor observation from some robot r'' that was previously transmitted to robot r will not be retained in the marginal distribution that robot r shares with any other robot r'. As a result, raw sensor observations from any robot are used at most once.

While not at the focus of this work, another solution might require that each robot r transmits, in addition to $p(X^r|Z^r)$, the measurement likelihood terms received from other robots as a means to propagate information in the system, thereby facilitating improved multirobot consistent estimation. Such an approach has its roots in [38] and has been recently applied to multirobot SLAM [37]. Additional approaches for consistent distributed estimation include [39] and [24].

Any robot r that receives an observation $z_{l'}^{r'}$ from some robot r', generates candidate correspondences by matching $z_{l'}^{r'}$ with its own informative observations $\{z_i^r\}$. Each such correct match $u_{l,l'}^{r,r'}$ between $z_{l'}^{r'}$ and $z_l^r \in \{z_i^r\}$ represents a (relative-pose) constraint involving the poses $x_{l'}^{r'}$ and x_l^r, with $l, l' \leq k$ where k is the current time index. Many of these constraints are outliers. The set $\mathcal{F}^r$ of multirobot data association, which is available to robot r at time t_k, includes all such individual multirobot data associations up to time t_k, $(r, r', l, l') \in \mathcal{F}_k^r$.

An example of the multirobot candidate correspondences set $\mathcal{F}^r$ is shown in Figure 2. The figure illustrates the candidate correspondences in $\mathcal{F}^r$ between the blue robot and other robots (green and red). Since the initial relative poses between the robots are unknown, these transformations were set to arbitrary values.

Given the multirobot data association $\mathcal{F}^r$, and the appropriate constraints $u_{l,l'}^{r,r'}$, the joint pdf $p(\mathcal{X}^r|\mathcal{Z}^r)$ from (3) can be expressed as

$$p(\mathcal{X}^r|\mathcal{Z}^r) \propto \prod_r p(X^r|Z^r) \prod_{(r,r',l,l')\in\mathcal{F}^r} p(u_{l,l'}^{r,r'}|x_l^r, x_{l'}^{r'}), \tag{8}$$

where $\mathcal{Z}^r$ includes both single-robot observations, for each robot, and all relative-pose multirobot constraints

$$\mathcal{Z}^r \doteq \cup_{r=1}^R Z^r \cup \{u_{l,l'}^{r,r'} | (r, r', l, l') \in \mathcal{F}^r\} \subseteq \mathcal{Z}^{\mathcal{R}}. \tag{9}$$

Because the robots express their local trajectories with respect to *different* reference systems, the measurement likelihood term in (8) is

$$p(u_{l,l'}^{r,r'}|x_l^r, x_{l'}^{r'}) \propto \exp\Big(-\frac{1}{2}\|\operatorname{err}(u_{l,l'}^{r,r'}, x_l^r, x_{l'}^{r'})\|_\Sigma^2\Big), \tag{10}$$

$$\operatorname{err}(u_{l,l'}^{r,r'}, x_l^r, x_{l'}^{r'}) \doteq u_{l,l'}^{r,r'} \ominus h(x_l^r, x_{l'}^{r'}), \tag{11}$$

$$h(x_l^r, x_{l'}^{r'}) \doteq x_l^r \ominus (T_{r'}^r \oplus x_{l'}^{r'}), \tag{12}$$

where $T_{r'}^r$ is a transformation between the reference frames of robots r and r', and the notation $\oplus$ represents the transformation compose operator [40]. Since the robots start operating from different unknown locations, this transformation is *initially unknown*.

Furthermore, while the formulation in (8) assumes multirobot data association $\mathcal{F}$ is given, in practice it is *unknown* ahead of time and must therefore be established. The next section focuses on simultaneously solving data association and the robot pose inference by introducing a latent variable.

Incorporating Multirobot Data Association

Instead of assuming multirobot data association is given, a latent binary variable $j_{l,l'}^{r,r'}$ is used to model inlier and outlier correspondences using the convention that correspondence $(r, r', l, l') \in \mathcal{F}^r$ is an inlier if $j_{l,l'}^{r,r'} = inlier$ and accordingly an outlier when $j_{l,l'}^{r,r'} = outlier$. Denoting all the latent variables representing data association between robot r and other

robots by $\mathcal{J}^r$, calculating the MAP inference over $\mathcal{X}^r$ corresponds to

$$\hat{\mathcal{X}}^r = \arg\max_{\mathcal{X}^r} p(\mathcal{X}^r | \mathcal{Z}^r) = \arg\max_{\mathcal{X}^r} \sum_{\mathcal{J}^r} p(\mathcal{X}^r, \mathcal{J}^r | \mathcal{Z}^r), \quad (13)$$

with

$$p(\mathcal{X}^r, \mathcal{J}^r | \mathcal{Z}^r) \propto \prod_r p(X^r | Z^r) \cdot \prod_{(r,r',l,l') \in \mathcal{F}^r} p(j_{l,l'}^{r,r'} | x_l^r, x_{l'}^{r'})\, p(u_{l,l'}^{r,r'} | x_l^r, x_{l'}^{r'}, j_{l,l'}^{r,r'}), \quad (14)$$

where the term $p(j_{l,l'}^{r,r'} | x_l^r, x_{l'}^{r'})$ is considered as the prior $p(j_{l,l'}^{r,r'})$, since it is not conditioned on any observations.

The likelihood term $p(u_{l,l'}^{r,r'} | x_l^r, x_{l'}^{r'}, j_{l,l'}^{r,r'})$ can be evaluated for both $j_{l,l'}^{r,r'} = inlier$ and $j_{l,l'}^{r,r'} = outlier$ with (10) using Σ_{in} and Σ_{out}, representing, respectively, the covariances corresponding to inlier and outlier distributions, with $\Sigma_{\text{in}} \ll \Sigma_{\text{out}}$.

However, when the initial relative pose $T_{r'}^r$ between any two robots r and r' is unknown, performing inference over (14) is doomed to fail [9]; since the transformation $T_{r'}^r$ is unknown and can only be arbitrarily set, each candidate multirobot data association $(r, r', l, l') \in \mathcal{F}^r$ with a corresponding constraint $u_{l,l'}^{r,r'}$ typically results in a high discrepancy between $\approx u_{l,l'}^{r,r'}$ and the prediction $h(x_l^r, x_{l'}^{r'})$ from (12), and therefore results in a high error $\text{err}(u_{l,l'}^{r,r'}, x_l^r, x_{l'}^{r'})$. These high errors are obtained both for inlier and outlier correspondences. Since $\Sigma_{\text{in}} \ll \Sigma_{\text{out}}$, the probability $p(u_{l,l'}^{r,r'} | x_l^r, x_{l'}^{r'}, j_{l,l'}^{r,r'})$ is higher for $j_{l,l'}^{r,r'} = outlier$ than $j_{l,l'}^{r,r'} = inlier$, regardless of whether the correspondence $(r, r', l, l') \in \mathcal{F}^r$ is an inlier or outlier in practice. As a result, attempting to infer both multirobot data association and robot trajectories leads to identifying all candidate correspondences in $\mathcal{F}^r$ as outliers.

It is for this reason that initial relative poses *must* be first estimated so that the error in (11) could be used to distinguish between inlier and outlier correspondences. The following sections describe an approach for concurrently estimating the transformations $T_{r'}^r$ and multirobot data association in an incremental distributed framework.

Remark

It may seem that outliers can be directly identified and rejected by matching algorithms such as RANSAC [31] or iterative closest point (ICP) [41]. However, this statement is only partially true. Although these algorithms are capable of accurate relative-pose estimation given sensor observations of a *common* scene, identifying the fact that two given observations were acquired from *different* parts of the environment is a much more challenging task. In the single-robot case, a related aspect is establishing loop closures. Current loop-closure techniques cannot guarantee outlier-free data associations, and approaches for robust graph optimization are still actively investigated. Similarly, in the multirobot framework, multirobot data association cannot be assumed outlier free, especially in the presence of perceptual aliasing (for example, when observing two corridors that look alike). These challenging scenarios are handled in the EM formulation and model-selection-based framework.

DISTRIBUTED INFERENCE OVER INITIAL RELATIVE POSE USING EM

The remaining formulation is based on the following key observation: given local robot trajectories, each candidate multirobot correspondence $(r, r', l, l') \in \mathcal{F}^r$, regardless of whether it is an inlier or outlier, suggests a solution for the transformation $T_{r'}^r$. However, only the inlier correspondences produce similar transformations, while those calculated from outlier correspondences typically disagree among each other, unless these outliers are caused by measurement aliasing.

This concept is illustrated in a synthetic toy example in Figure 5. In the figure, ground-truth robot trajectories are given in (a), and candidate matches between two of the robots (red and blue) are shown in (b), with robot initial poses set to

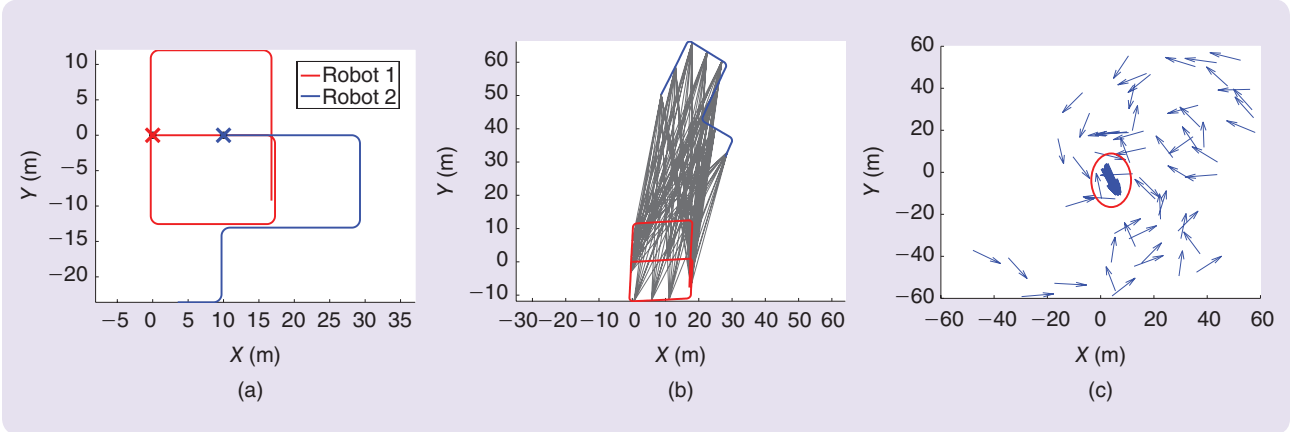


FIGURE 5 A simulation experiment depicting (a) robot trajectories. Initial positions are denoted by cross marks. (b) Correspondences between trajectories, and (c) distribution of relative pose transformations. The large cluster marked in red circle in (c) corresponds to the correct transformation.

arbitrary values. Many of these correspondences are outliers—this simple example assumes there is no measurement aliasing. The distribution of the transformation $T_r^{r'}$, calculated for each correspondence, is shown in Figure 5(c). The cluster corresponding to the correct transformation is larger than the rest. Real-world scenarios, however, often exhibit some level of measurement aliasing, which often leads to multiple clusters and further complicates the identification of the correct transformation $T_r^{r'}$. Examples of such scenarios are given throughout this article and the approach developed to address this challenge is discussed below.

Based on the above key observation, an EM optimization [42] is formulated that allows each robot to recover the initial relative pose with respect to other robots (and therefore establish a common reference frame) in a distributed manner.

The MAP estimate of an initial relative pose between robots r and r' can be written as

$$\hat{T}_r^{r'} = \arg\max_{T_r^{r'}} p\left(T_r^{r'} \middle| \hat{\mathcal{X}}^r, \mathcal{Z}^r\right), \tag{15}$$

where $\hat{\mathcal{X}}^r$ and $\mathcal{Z}^r$ are defined in (7) and (9). This inference can be solved by each robot r in a distributed fashion, based on the available observations $\mathcal{Z}^r$.

Similar to (13), latent variables $\mathcal{J}^r$ are included in this inference problem

$$\hat{T}_r^{r'} = \arg\max_{T_r^{r'}} \sum_{\mathcal{J}^r} p\left(T_r^{r'}, \mathcal{J}^r \middle| \hat{\mathcal{X}}^r, \mathcal{Z}^r\right). \tag{16}$$

However, since the above inference involves accounting for all the possible values for each of the latent binary variables in $\mathcal{J}^r$, it is computationally intractable.

In contrast, the EM approach enables a drastic reduction in computational complexity while producing a locally optimal solution. The tth iteration of the EM framework is (see "Expectation Maximization Equations" for full details)

Expectation Maximization Equations

The following details the underlying expectation maximization (EM) optimization equations for recovering a local maxima of the initial relative pose $T_{r'}^{r}$ using (17). Recall the MAP estimate of $T_{r'}^{r}$ in (16)

$$\hat{T}_{r'}^{r} = \arg\max_{T_{r'}^{r}} \sum_{\mathcal{J}^r} p\left(T_{r'}^{r}, \mathcal{J}^r \middle| \hat{\mathcal{X}}^r, \mathcal{Z}^r\right).$$

To simplify notation, the superscripts and subscripts r and r' are removed, and this equation is rewritten as

$$\hat{T} = \arg\max_{T} \sum_{\mathcal{J}} p\left(T, \mathcal{J} \middle| \hat{\mathcal{X}}, \mathcal{Z}\right). \tag{S9}$$

The tth iteration of the EM optimization can be described by the following two steps [42].

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

$$\begin{aligned} Q(T \,|\, \hat{T}^{(i)}) &\doteq \mathbb{E}_{\mathcal{J} | \hat{T}^{(i)}, \hat{\mathcal{X}}, \mathcal{Z}}\left[\log p\left(T, \mathcal{J} \middle| \hat{\mathcal{X}}, \mathcal{Z}\right)\right] \\ &= \int p\left(\mathcal{J} \middle| \hat{T}^{(i)}, \hat{\mathcal{X}}, \mathcal{Z}\right) \log p\left(T, \mathcal{J} \middle| \hat{\mathcal{X}}, \mathcal{Z}\right) d\mathcal{J}. \end{aligned} \tag{S10}$$

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

$$\hat{T}^{(i+1)} = \arg\max_{T} Q(T \,|\, \hat{T}^{(i)}).$$

Focusing on the E step, the joint pdf $p\left(T, \mathcal{J} \middle| \hat{\mathcal{X}}, \mathcal{Z}\right)$ is

$$p\left(T, \mathcal{J} \middle| \hat{\mathcal{X}}, \mathcal{Z}\right) \propto p\left(\mathcal{Z} \middle| \hat{\mathcal{X}}, \mathcal{J}, T\right) p\left(\mathcal{J}, T \middle| \hat{\mathcal{X}}\right).$$

Assuming $\mathcal{J}$ and T are independent given robot poses $p\left(\mathcal{J}, T \middle| \hat{\mathcal{X}}\right) = p(T \,|\, \hat{\mathcal{X}}) p\left(\mathcal{J} \middle| \hat{\mathcal{X}}\right)$, the above can be expressed in terms of individual candidate correspondences [see (14)]

$$p\left(\mathcal{J}, T \middle| \hat{\mathcal{X}}\right) \propto p(T \,|\, \hat{\mathcal{X}}^r) \prod_{(r,r',l,l') \in \mathcal{F}} p(j_{l,l'}^{r,r'} \,|\, \hat{x}_l^r, \hat{x}_{l'}^{r'}) p(u_{l,l'}^{r,r'} \,|\, x_l^r, x_{l'}^{r'}, j_{l,l'}^{r,r'}, T).$$

Assuming the binary variables $j_{l,l'}^{r,r'} \in \mathcal{J}$ are statistically independent conditioned on $\hat{T}^{(t)}, \hat{\mathcal{X}}$, and $\mathcal{Z}$, (S10) can be written as

$$\begin{aligned} Q(T \,|\, \hat{T}^{(i)}) = \log p(T \,|\, \hat{\mathcal{X}}^r) + \sum_{s \in \mathcal{F}^r} \int p\left(j_s \middle| \hat{T}^{(i)}, \hat{x}_l^r, \hat{x}_{l'}^{r'}, u_{l,l'}^{r,r'}\right) \\ \cdot \log[p(j_s \,|\, \hat{x}_l^r, \hat{x}_{l'}^{r'}) p(u_{l,l'}^{r,r'} \,|\, \hat{x}_l^r, \hat{x}_{l'}^{r'}, j_s, T)] dj_s, \end{aligned} \tag{S11}$$

where $s \doteq (r, r', l, l') \in \mathcal{F}$ and $j_s \doteq j_{l,l'}^{r,r'}$, and where

$$p\left(j_s \middle| \hat{T}^{(i)}, \hat{x}_l^r, \hat{x}_{l'}^{r'}, u_{l,l'}^{r,r'}\right) = \frac{p\left(u_{l,l'}^{r,r'} \middle| \hat{x}_l^r, \hat{x}_{l'}^{r'}, j_s, \hat{T}^{(t)}\right) p(j_s \,|\, \hat{x}_l^r, \hat{x}_{l'}^{r'}, \hat{T}^{(t)})}{p\left(u_{l,l'}^{r,r'} \middle| \hat{T}^{(t)}, \hat{x}_l^r, \hat{x}_{l'}^{r'}\right)}. \tag{S12}$$

The term $p(j_s \,|\, \hat{x}_l^r, \hat{x}_{l'}^{r'}, \hat{T}^{(t)})$ in the above equation, and $p(j_{l,l'}^{r,r'} \,|\, \hat{x}_l^r, \hat{x}_{l'}^{r'})$ in (S11), are modeled by the prior $p(j_s)$, since both are not conditioned on any observations. The measurement likelihood term $p\left(u_{l,l'}^{r,r'} \middle| \hat{x}_l^r, \hat{x}_{l'}^{r'}, j_s, \hat{T}^{(t)}\right)$ in (S12), which is similar to (10), is

$$p\left(u_{l,l'}^{r,r'} \middle| \hat{x}_l^r, \hat{x}_{l'}^{r'}, j_s, \hat{T}^{(t)}\right) \propto \exp\left(-\frac{1}{2} \| \operatorname{err}(u_{l,l'}^{r,r'}, \hat{x}_l^r, \hat{x}_{l'}^{r'}) \|_{\Sigma_{j_s}}^2\right), \tag{S13}$$

where the function err(.) is (11) and $\Sigma_{j_s} \in \{\Sigma_{\text{inlier}}, \Sigma_{\text{outlier}}\}$.

To conclude, note that the denominator in (S12) is not actually evaluated; instead, only the numerator is calculated for both inlier and outlier models, and then normalized so that

$$p\left(j_s = \text{inlier} \middle| \hat{T}^{(i)}, \hat{x}_l^r, \hat{x}_{l'}^{r'}, u_{l,l'}^{r,r'}\right) + p\left(j_s = \text{outier} \middle| \hat{T}^{(i)}, \hat{x}_l^r, \hat{x}_{l'}^{r'}, u_{l,l'}^{r,r'}\right) = 1.$$

$$\hat{T}^{(t)} = \arg\max_{T} \mathbb{E}_{\mathcal{J}^r|\hat{T}^{(t-1)},\hat{\mathcal{X}}^r,\mathcal{Z}^r} \log\left[p\left(T, \mathcal{J}^r \middle| \hat{\mathcal{X}}^r, \mathcal{Z}^r\right)\right]$$
$$= \arg\max_{T} \sum_{\mathcal{J}^r} p\left(\mathcal{J}^r \middle| \hat{T}^{(t-1)}, \hat{\mathcal{X}}^r, \mathcal{Z}^r\right) \log\left[p\left(T, \mathcal{J}^r \middle| \hat{\mathcal{X}}^r, \mathcal{Z}^r\right)\right], \tag{17}$$

where the notation $T \doteq (T_{r'}^r)^{(t)}$ is used for brevity.

Solving (17) involves a nonlinear optimization that converges to a *local* minima given an initial guess for a solution. The optimization problem in (17) is highly nonconvex, so an initial solution in the vicinity of the global minima is required to guarantee convergence to the *correct* transformation $T_{r'}^r$.

Recalling the key observation from the beginning of this section (inlier correspondences produce similar transformations, and outlier correspondences yield inconsistent transformations) and using the available robot local trajectories $\hat{\mathcal{X}}^r$, an initial relative pose is calculated for each correspondence $(r, r', l, l') \in \mathcal{F}^r$, resulting in a set $\mathcal{T}_{r'}^r(\mathcal{F}^r)$. The distribution of this set is then analyzed to perform a naïve clustering for each element in $T_{r'}^r$ separately. The k most common values that are sufficiently far away from each other are chosen. Instead of considering all the possible combinations for each axis as an initial solution (k^n where n is the dimensionality of $T_{r'}^r$), any entries that do not have a nearby initial relative pose in the set $\mathcal{T}_{r'}^r(\mathcal{F}^r)$ are discarded. The actual choice of the parameter k depends on the available computational resources, since this parameter determines the number of times the EM optimization is to be performed. In implementation, $k = 5$ is chosen, and two entries are considered to be sufficiently apart if the difference is above the corresponding standard deviation in Σ_{in} is 0.5 m for position and 0.05 rad for rotation.

This procedure generates several initial solutions for $T_{r'}^r$; running the EM optimization in (17) on each one of these solutions produces locally optimal initial relative poses $\hat{T}_{r'}^r$. Different hypotheses that converge to the same value $\hat{T}_{r'}^r$ are merged. Each of these estimates is optimal given the corresponding partitioning of the multirobot data association $\hat{\mathcal{J}}^r$ into inliers I and outliers O, calculated in the E step in EM; see (17).

The hypothesis $h \doteq \{I, O\}$ can now be used in conjunction with $I \cup O = \mathcal{J}^r$ to represent each such partition and collect all these hypotheses into the set $\mathcal{H}$. Note that each hypothesis $h \in \mathcal{H}$ leads to an estimate of $T_{r'}^r$ according to

$$\hat{T}_{r'}^r(h) = \arg\max_{T_{r'}^r} p\left(T_{r'}^r \middle| h, \hat{\mathcal{X}}^r, \mathcal{Z}^r\right). \tag{18}$$

Figure 3(b) illustrates the distribution of relative-pose constraints for each correspondence $(r, r', l, l') \in \mathcal{F}^r$. (The planar case is considered; the x and y axes correspond to position and the blue arrow direction represents orientation.) The transformations $\hat{T}_{r'}^r(h)$ for each hypothesis $h \in \mathcal{H}$ are denoted by red circles.

Often there is more than one hypothesis in $\mathcal{H}$, and a decision must be made about which hypothesis $h^* \in \mathcal{H}$. The chosen hypothesis will determine the initial relative pose, $\hat{T}_{r'}^r = \hat{T}_{r'}^r(h^*)$, which is used later to perform inference over robot trajectories. Therefore, identifying the *correct* hypothesis in $\mathcal{H}$ is crucial, especially in the presence of perceptual aliasing. The next section develops a probabilistic method for choosing the most probable hypothesis, focusing in particular on the complexities arising from the incremental framework that is essential for real-time operation.

HYPOTHESIS MODEL-BASED SELECTION

This section develops a probabilistic approach for choosing the most likely hypothesis, given the current information from the set $\mathcal{H}$. An incremental setting, in which information, such as robot trajectories, local sensor observations, and multirobot correspondences, is obtained gradually, adds several challenges. A criterion is needed to determine whether a sufficient amount of information has been accumulated to reliably estimate the initial relative poses. For example, robot trajectories and observed environments may initially not overlap (or not overlap at all), in which case all multirobot correspondences are outliers.

Measurement aliasing, meaning observations of similar environments, complicate the problem even further. Consider the robots start operating in two nearby somewhat *similar* environments (for example, corridors). Each robot r shares informative sensor observations, matches observations transmitted from other robots against its own informative observations, and adds the matches as candidate correspondences into the set $\mathcal{F}^r$. Since the two environments are similar but different, these matches are not only outliers but are also consistent with each other, which suggests that the two environments are actually the same. This consistency translates directly into a hypothesis $h = \{I, O\}$ with many inliers, all of which are actually (consistent) outliers. Note that even after the robots observe some areas in common, the correct hypothesis (corresponding to true inliers) must compete with the consistent-outlier hypothesis. Making a wrong decision and choosing an incorrect hypothesis can lead to catastrophic results because outliers are included in the inference in (3).

Figure 3 shows such a case in a real-world scenario. The robots travel in three different corridors, as shown in Figure 3(a). However, because these corridors are similar in appearance, many consistent matches between scans from the three corridors are generated, leading to the clusters 1–4 emphasized by red ellipses in Figure 3(b). Making a wrong decision and choosing clusters 2–4 results in incorrect alignment of robot trajectories and all the true inlier correspondences are identified as outliers.

This challenging problem is approached within a model-selection framework, which aims to calculate the probability of each hypothesis h in the set $\mathcal{H}$: $p\left(h \middle| \hat{\mathcal{X}}^r, \mathcal{Z}^r\right)$. These

This article develops a distributed and incremental multirobot approach that allows a group of robots to concurrently establish a common reference frame and resolve multirobot data association on the fly.

probabilities can be obtained by integrating over all possible values of the continuous random variable $T^r_{r'}$

$$p\left(h \middle| \hat{\mathcal{X}}^r, \mathcal{Z}^r\right) = \int_{T^r_{r'}} p\left(T^r_{r'}, h \middle| \hat{\mathcal{X}}^r, \mathcal{Z}^r\right). \tag{19}$$

Applying Bayes' rule,

$$p\left(h \middle| \hat{\mathcal{X}}^r, \mathcal{Z}^r\right) = \int_{T^r_{r'}} c \cdot p\left(\mathcal{Z}^r \middle| T^r_{r'}, h, \hat{\mathcal{X}}^r\right) p\left(T^r_{r'}, h \middle| \hat{\mathcal{X}}^r\right), \tag{20}$$

with a hypothesis-independent constant $c \doteq 1/p\left(\mathcal{Z}^r \middle| \hat{\mathcal{X}}^r\right)$ that need not be calculated. Factoring the last term as

$$p\left(T^r_{r'}, h \middle| \hat{\mathcal{X}}^r\right) = p\left(T^r_{r'} \middle| h, \hat{\mathcal{X}}^r\right) p\left(h \middle| \hat{\mathcal{X}}^r\right), \tag{21}$$

(20) turns into

$$p\left(h \middle| \mathcal{Z}^r, \hat{\mathcal{X}}^r\right) = c \cdot p\left(h \middle| \hat{\mathcal{X}}^r\right) p\left(\mathcal{Z}^r \middle| h, \hat{\mathcal{X}}^r\right), \tag{22}$$

with

$$p\left(\mathcal{Z}^r \middle| h, \hat{\mathcal{X}}^r\right) = \int_{T^r_{r'}} p\left(T^r_{r'} \middle| h, \hat{\mathcal{X}}^r\right) p\left(\mathcal{Z}^r \middle| T^r_{r'}, h, \hat{\mathcal{X}}^r\right). \tag{23}$$

To calculate hypothesis probability in (20), each of the terms in (22) and (23) needs to be evaluated: the hypothesis prior $p\left(h \middle| \hat{\mathcal{X}}^r\right)$, relative-pose transformation prior $p\left(T^r_{r'} \middle| h, \hat{\mathcal{X}}^r\right)$, and the measurement likelihood term $p\left(\mathcal{Z}^r \middle| T^r_{r'}, h, \hat{\mathcal{X}}^r\right)$.

Before presenting the equations for each of these terms, it is beneficial to first discuss the importance of the hypothesis prior term in deciding which hypothesis to choose from the set $\mathcal{H}$. As will be seen below, the measurement likelihood $p\left(\mathcal{Z}^r \middle| T^r_{r'}, h, \hat{\mathcal{X}}^r\right)$ essentially prioritizes hypotheses with more inliers. However, it does not address the question of whether a sufficient amount of information has been accumulated to make a decision. In other words, given a set $\mathcal{H}$, there is always a hypothesis with the highest measurement likelihood. A criterion is needed for whether the hypothesis is unambiguous and should be chosen.

This aspect is particularly crucial in the incremental setting in the context of measurement aliasing, which can lead to a dominant hypothesis of consistent outliers. Note that each correspondence, regardless of whether it is considered to be an inlier or an outlier by a hypothesis, is the result of a high-quality match between two sensor observations (for example, an ICP match between two laser scans). Therefore, the main reason for a correspondence to be an outlier in practice (in other words, a highly confident match between two scans from different areas) is perceptual aliasing. Relying only on the measurement likelihood term, it is easy to mistakenly choose this incorrect hypothesis.

This critical issue is addressed by arguing that the hypothesis prior $p\left(h \middle| \hat{\mathcal{X}}^r\right)$ can provide insight as to how likely the hypothesis $h = \{I, O\}$ is in the first place. An underlying basic assumption in the approach is that the robots must observe common places eventually, not necessarily at the same time. This assumption makes sense for indoor navigation, for example, since the robots are operating in closed and finite environments. However, the assumption is often also reasonable for collaborative navigation in outdoor environments, as demonstrated experimentally below.

Having gained a conceptual understanding for the importance of the hypothesis prior term, the following sections proceed with the equation derivations for each of the mentioned probability terms in (22) and (23).

Measurement Likelihood $p\left(\mathcal{Z}^r \middle| T^r_{r'}, h, \hat{\mathcal{X}}^r\right)$

The measurement likelihood is discussed first. Recalling the hypothesis definition, $h = \{I, O\}$, and writing individual inlier and outlier measurement terms gives

$$p\left(\mathcal{Z}^r \middle| T^r_{r'}, h, \hat{\mathcal{X}}^r\right) = k(h) \prod_{i \in I} \exp\left(-\frac{1}{2} \left\| \operatorname{err}\left(u_i, \hat{\mathcal{X}}^r, T^r_{r'}\right) \right\|^2_{\Sigma_{\text{in}}}\right) \cdot \prod_{o \in O} \exp\left(-\frac{1}{2} \left\| \operatorname{err}\left(u_o, \hat{\mathcal{X}}^r, T^r_{r'}\right) \right\|^2_{\Sigma_{\text{out}}}\right), \tag{24}$$

where u_i and u_o are used to represent the appropriate relative-pose constraints $u^{r,r'}_{I,I'}$, and $err(\,.\,)$ is defined in (11). The hypothesis-dependent coefficient $k(h)$ is

$$k(h) \doteq \prod_{i \in I} \frac{1}{\sqrt{|2\pi\Sigma_{\text{in}}|}} \prod_{o \in O} \frac{1}{\sqrt{|2\pi\Sigma_{\text{out}}|}}. \tag{25}$$

Considering $p\left(T^r_{r'} \middle| h, \hat{\mathcal{X}}^r\right)$ can be expressed as the Gaussian

$$p\left(T^r_{r'} \middle| h, \hat{\mathcal{X}}^r\right) = N(T_0, \Sigma_0), \tag{26}$$

with known parameters T_0 and Σ_0 (as discussed in initial relative-pose prior section), the integrand $q(T^r_{r'})$ in (23),

$$q(T^r_{r'}) \doteq p\left(T^r_{r'} \middle| h, \hat{\mathcal{X}}^r\right) p\left(Z \middle| T^r_{r'}, h, \hat{\mathcal{X}}^r\right), \tag{27}$$

can be expressed using (24)–(26). Since the involved distributions are all Gaussians, this expression can be represented by a single Gaussian

$$q(T^r_{r'}) = N(\hat{T}^r_{r'}, \Sigma_{\text{MAP}}), \tag{28}$$

where the MAP estimate is obtained by $\hat{T}^r_{r'} = \arg\max_{T^r_{r'}} q(T^r_{r'})$ and can be then used as a linearization point to calculate the covariance Σ_{MAP}.

Recall that

$$\int_x a \exp(\|x - \hat{x}\|^2_{\Sigma_{\text{MAP}}}) = a\sqrt{|2\pi\Sigma_{\text{MAP}}|}, \tag{29}$$

where the notations $\hat{x}$ and Σ_{MAP} represent, respectively, the MAP estimate and covariance of some random variable x. Substituting (27) and (28) into (22) and using (29) gives an expression for the hypothesis probability

$$p\left(h \middle| \mathcal{Z}^r, \hat{\mathcal{X}}^r\right) = c \cdot p(h|\hat{\mathcal{X}}^r) k'(h) \sqrt{|2\pi\Sigma_{\text{MAP}}|}, \tag{30}$$

where $k'(h) \doteq k(h) \cdot 1/\sqrt{|2\pi\Sigma_0|}$. The priors $p(h|\hat{\mathcal{X}}^r)$ and $p\left(T^r_{r'} \middle| h, \hat{\mathcal{X}}^r\right)$ are discussed next.

Hypothesis Prior $p(h|\hat{\mathcal{X}}^r)$

As previously mentioned, the hypothesis prior is an important component in deciding which hypothesis $h = \{I, O\}$ to choose from the set $\mathcal{H}$ while enhancing robustness to perceptual aliasing. Given the robot local trajectories, the set $\mathcal{F}^r$ can be partitioned into inlier and outlier sets I and O, respectively, even without the actual measurements of the corresponding constraints. Further, an additional hypothesis $h_0 = \{I_0, O_0\}$ is added to the set $\mathcal{H}$: this hypothesis (referred to as the *null hypothesis*) corresponds to perceptual aliasing. The null hypothesis represents the possibility that all of the correspondences are actually outliers ($I_0 \equiv \phi$).

The prior probabilities of all hypotheses in $\mathcal{H}$, including the null hypothesis h_0, can now be calculated. As further detailed below, the most likely hypothesis h, according to the posterior $p\left(h \middle| \hat{\mathcal{X}}^r, \mathcal{Z}^r\right)$ from (30) is then chosen only if its prior probability $p(h|\hat{\mathcal{X}}^r)$ is significantly dominant compared to all other hypotheses in $\mathcal{H}$. In this way, the algorithm waits until sufficient information is accumulated to disambiguate which hypothesis is the correct one.

A previously mentioned, the underlying basic assumption is that the robots must eventually observe common places (but not necessarily concurrently). Each given candidate correspondence can therefore represent the same place, observed by two different robots or two different places. However, the number of unique places is unknown ahead of time. The CRP (and closely related Dirichlet process), which has previously been used for topological mapping [43], models this problem probabilistically [12], [13].

The model in (S8) is now used to develop an approach for quantifying the hypothesis prior for each $h \in \mathcal{H}$. The approach is based on the following interpretation of inlier and outlier correspondences determined by a given hypothesis $h = \{I, O\} \in \mathcal{H}$. An inlier $i \in I$ relating two different poses of robots r and r' indicates the two poses are close to each other (generating overlapping observations). Accordingly, the involved two poses in an outlier correspondence $o \in O$ are considered to be two different (unique) places.

More formally, let m_{in} and m_{out} represent, respectively, the number of inliers and outliers in a given hypothesis $h = \{I, O\} \in \mathcal{H}$ and denote by m the overall number of candidate correspondences. Note that m is the same for all hypotheses in $\mathcal{H}$, and also $m = m_{\text{in}} + m_{\text{out}}$.

Given some hypothesis $h = \{I, O\} \in \mathcal{H}$, the number of unique places due to inlier correspondences is approximately captured in m_{in}. In other words, each inlier correspondence is modeled to contribute a single unique place: as mentioned before, a new informative sensor observation is shared by each robot r only if it has moved certain distance (0.5 m when operating indoors) since it shared the previous informative observation. Therefore, any two inlier correspondences $i_1, i_2 \in I$, contain some unique information about the environment. Moreover, only a small portion of the observation pairs (originating from different robots) are correctly matched and added as candidate correspondences. Consequently, the number of unique observed places due to inlier correspondences is modeled as m_{in}.

Similarly, any two pairs of outlier correspondences $o_1, o_2 \in O$ are considered to represent four different places, thus approximately capturing the number of unique observed places due to outliers as $2m_{\text{out}}$. The unique places contributed by inliers and outliers are assumed to not overlap.

Using the above, the number of unique places for each given hypothesis is modeled as

$$n(h) = 2m_{\text{out}}(h) + m_{\text{in}}(h), \tag{31}$$

where the parameters m_{out} and m_{in} are hypothesis specific. Observe that while the overall number of candidate correspondences m is the *same* for all hypotheses in $\mathcal{H}$, the number of unique places n changes from one hypothesis to another in the range $n \in [m, 2m]$. In particular, n always assumes the highest value for the null hypothesis ($m_{\text{out}} = m, m_{\text{in}} = 0$) and, on the other extreme, the lowest value for the all-inliers hypothesis ($m_{\text{out}} = 0, m_{\text{in}} = m$), if such a hypothesis exists in $\mathcal{H}$.

The prior probability $p(h|\hat{\mathcal{X}}^r)$ of each hypothesis $h \in \mathcal{H}$ can now be calculated from (S8) using the hypothesis-specific value for n that is calculated according to (31). The prior probabilities of all hypotheses in $\mathcal{H}$ are normalized by

$$\eta \doteq \sum_{h \in \mathcal{H}} f(n(h)) \tag{32}$$

to obtain a valid probability distribution

$$p(h|\hat{\mathcal{X}}^r) = \frac{f(n(h))}{\eta}. \tag{33}$$

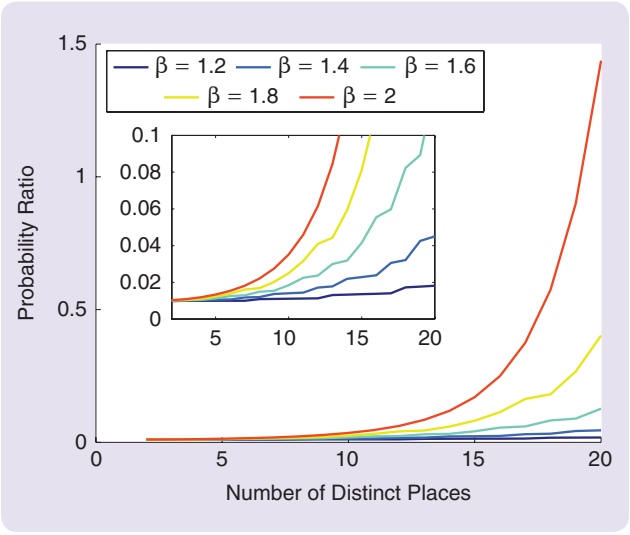


FIGURE 6 The probability ratio $f(n)/f(\beta n)$ with $1 \le \beta \le 2$.

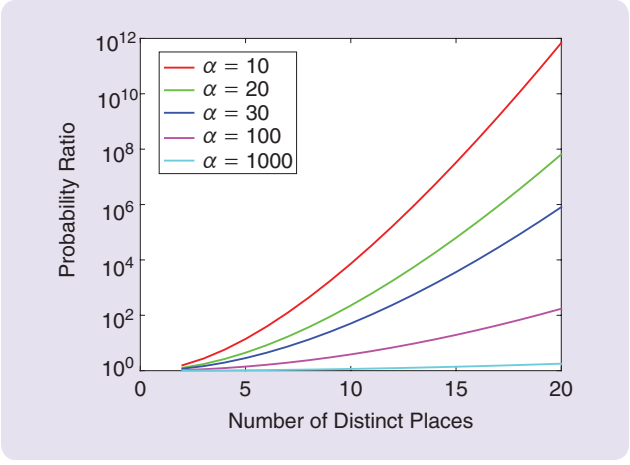


FIGURE 7 The probability ratio $f(n)/f(\beta n)$ for different values of α, with $\beta = 2$.

Equation (33) provides a mechanism to decide if sufficient information has been accumulated to indicate that a certain hypothesis is dominant with respect to competing hypotheses while explicitly accounting for potential measurement aliasing by introducing the null hypothesis.

In particular, for a given m and α, hypotheses with many inliers span different parts of robot trajectories. These hypotheses receive higher probabilities since the number of unique places n is smaller for such hypotheses. The two extreme cases are hypotheses with only inlier or only outlier correspondences, corresponding, respectively, to $n = m$ and $n = 2m$. In general, the parameter n for some hypothesis is γm, with $1 \le \gamma \le 2$. It can be verified from (S8) that $f(\gamma m)$ decreases with γ [see also Figure S2(b)]. The intuition here is that, at each given time (fixed m), a hypothesis whose inliers involve larger portions of robot trajectories (smaller n) is less prone to measurement aliasing over a hypothesis whose inliers only involve a small area (larger n).

A decision must eventually be made regarding whether sufficient information has been accumulated to choose a hypothesis. To address this decision, the CRP model in (S8) can be recalled to formulate the following lemma.

Lemma 1

The ratio between hypotheses prior probabilities increases as more information is accumulated (or, as more unique places are observed).

Proof

The number of unique places in any two hypotheses $h_i, h_j \in \mathcal{H}$ can be denoted by n_i and n_j, with $n_i, n_j \in [m, 2m]$. Without loss of generality, assuming $n_i < n_j$ implies $p(h_i|\hat{\mathcal{X}}^r) > p(h_j|\hat{\mathcal{X}}^r)$. Since both probabilities are calculated using the same normalization constant η in (32), $f(n_i) > f(n_j)$.

The parameter $\beta \in [1,2]$ may always be found such that $n_j = \beta n_i$, so the ratio $f(n_i)/f(n_j) \equiv f(n_i)/f(\beta n_i)$ can be expressed with (S8) as

$$\begin{aligned}\frac{f(n_i)}{f(\beta n_i)} &\doteq \frac{\alpha^{n_i}}{\prod_{j=1}^{n_i}(\alpha + j - 1)} \cdot \frac{\prod_{j=1}^{\beta n_i}(\alpha + j - 1)}{\alpha^{\beta n_i}} \\ &= \frac{1}{\alpha^{(\beta-1)n_i}} \prod_{j=n_i+1}^{\beta n_i} (\alpha + j - 1) \\ &= \frac{1}{\alpha^{(\beta-1)n_i}} \frac{(\alpha + \beta n_i - 1)!}{(\alpha + n_i)!}. \end{aligned} \tag{34}$$

Since $\beta \in [1,2]$ and $\alpha > 1$, (34) monotonically increases with n_i, as shown in Figure 6. However, increasing n_i corresponds to accumulating more information and adding new candidate correspondences to the set $\mathcal{F}$. Therefore, with time, the prior for any two given hypotheses becomes more distinguishable. ■

The implication of Lemma 1 is that, as more information is accumulated, it becomes possible to disambiguate between the different hypotheses. The parameter α determines how fast this process is, and therefore this parameter can be considered a tuning parameter to be set according to prior knowledge regarding the expected environment size, if such knowledge exists. In particular, also if α is too small, the null hypothesis is prematurely down weighted, which thereby increases sensitivity to measurement aliasing. On the other hand, larger values of α require an accumulation of more information before the prior of the appropriate hypothesis becomes sufficiently dominant to facilitate reliable decision. Figure 7 provides further insight regarding the influence of the parameter α on the ratio between hypotheses prior probabilities in (34), for $\beta = 2$. Noting the logarithmic scale of y-axis of the plot, it can be seen that the probability ratio significantly varies for different values of α. The gap between two competing hypotheses therefore increases as more information is accumulated, at different rates for different values of α. The development of

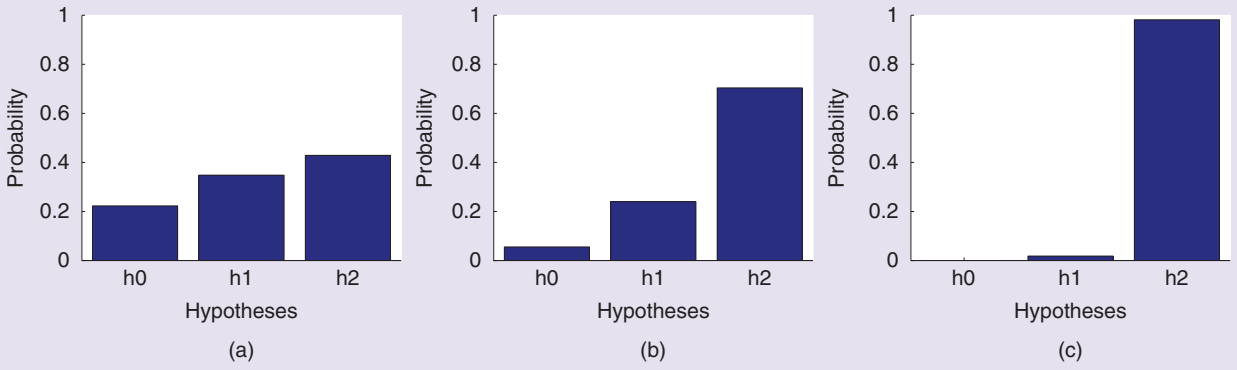


FIGURE 8 The prior probabilities of three different hypotheses h_0, h_1, and h_2 (see Table 2) as a function of m. A larger m corresponds to the accumulation of additional information. Parameter α is set to 50 in all cases. (a) $m = 5$. (b) $m = 10$. (c) $m = 20$.

specific approaches to determine α from available statistical information is outside the scope of this article and is left to future research.

This concept is illustrated by a toy example in Figure 8 that considers three different hypotheses that are detailed in Table 2: the null hypothesis (h_0), the hypothesis with half of the correspondences inliers (h_1), and the hypothesis with only inliers (h_2). Figure 8 shows the evolution of the hypotheses priors as a function of m. The prior

TABLE 2 A toy example with three hypotheses to illustrate the prior hypothesis.

h	Description	m_{in}	m_{out}	n
h_0	Null hypothesis	0	m	$2m$
h_1	Half inliers, half outliers	$0.5m$	$0.5m$	$1.5m$
h_2	All inliers	m	0	m

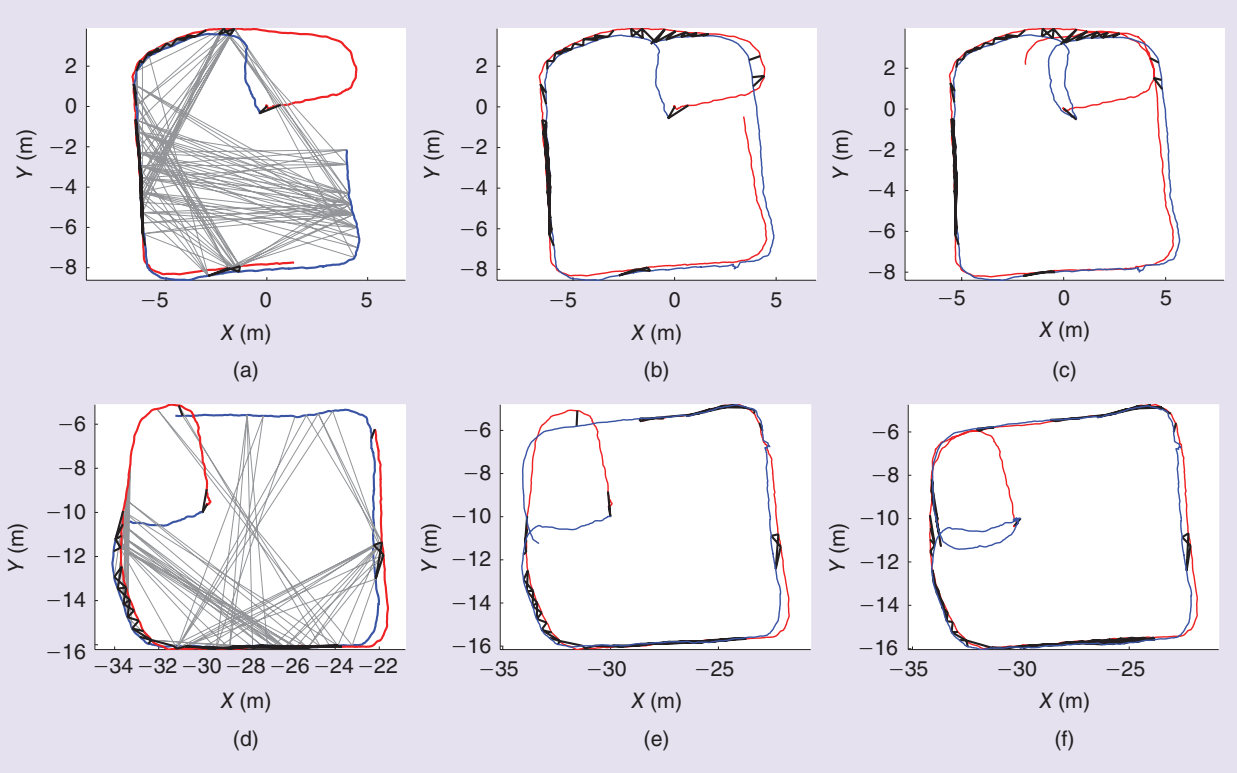


FIGURE 9 Robots' estimates of each other's trajectories expressed in local frame of each robot over time. (a)–(c) The red robot and (d)–(f) the blue robot. Parts (a) and (d) show the identified inlier (black) and outlier (gray) correspondences after a common reference frame has been established.

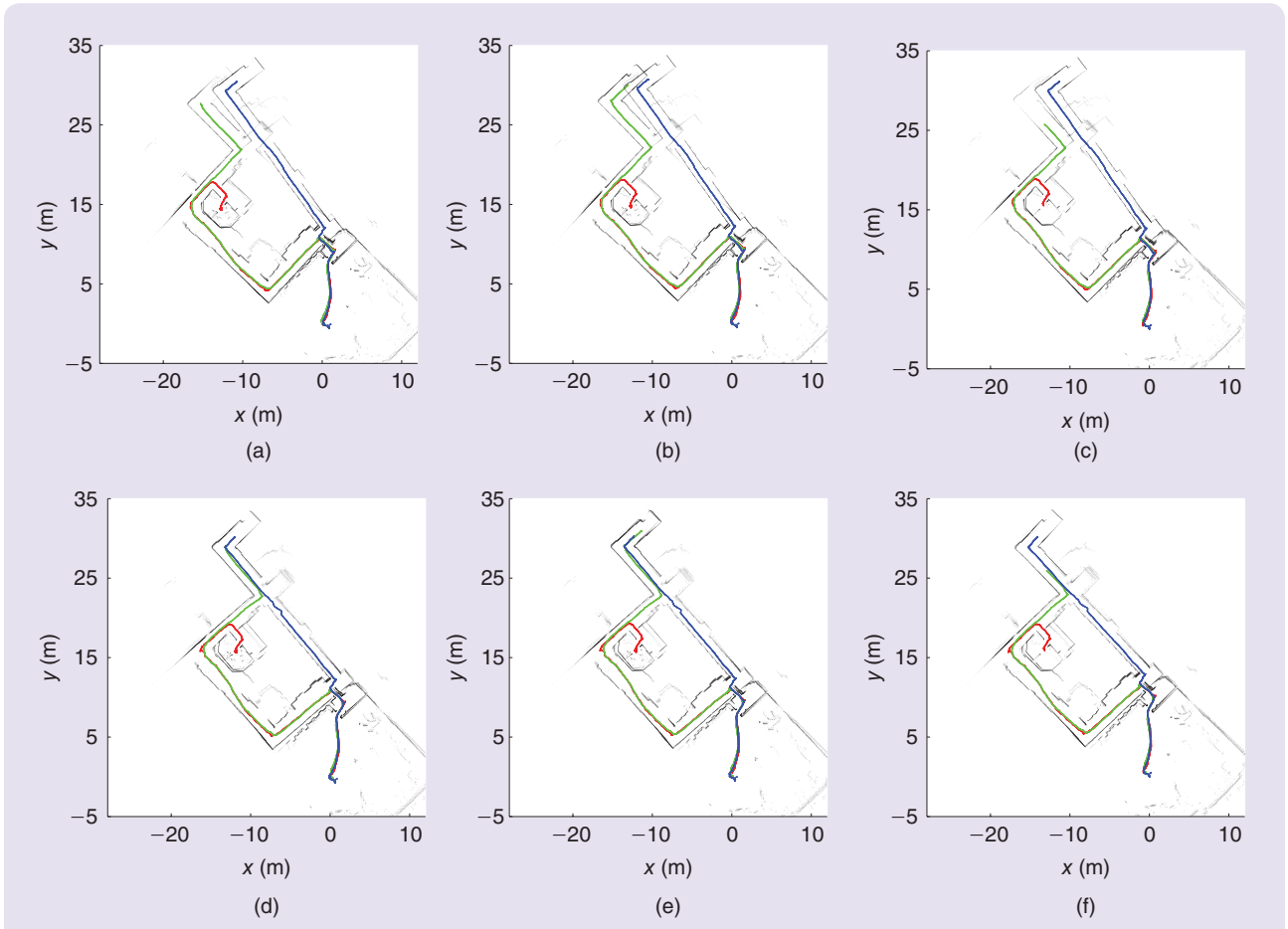


FIGURE 10 Data set **D2**: (a)–(c) the final estimated trajectories of all three robots from individual robot perspectives, without considering odometric uncertainty; (d)–(f) the estimated trajectories after accounting for odometric uncertainty. (a) By the red robot, (b) by the green robot, (c) by the blue robot, (d) by the red robot, (e) by the green robot, and (f) by the blue robot.

probability of the null hypothesis decreases as m increases, and the prior for h_2 (all inliers) gradually increases. The gap between the latter and the other two hypotheses (h_0 and h_1) drastically increases as well, corresponding to higher confidence level in choosing h_2 as a valid hypothesis candidate.

Initial Relative-Pose Prior $p\left(T_{r'}^{r} \middle| h, \hat{\mathcal{X}}^{r}\right)$

Given the inlier and outlier correspondences in $\mathcal{F}^r$ but without the actual measurements, the prior on the initial relative pose between the two robots $p\left(T_{r'}^{r} \middle| h, \hat{\mathcal{X}}^{r}\right)$ is uninformative.

Although, in principle, values of $T_{r'}^{r}$ could be analyzed more carefully to deem which are unreasonable, a basic approach is employed in this work. Because of the way constraints are generated, it is assumed that, on average, the relative-pose measurements $u_{l,l'}^{r,r'}$ are zero biased and therefore each measurement likelihood term can be considered to be distributed according to $N(0, \Sigma_{\text{high}})$ where Σ_{high} is a high uncertainty covariance (for example, 10 m in position and 90° in rotation). $p\left(T_{r'}^{r} \middle| h, \hat{\mathcal{X}}^{r}\right)$ is therefore modeled as the Gaussian

$$p\left(T_{r'}^{r} \middle| h, \hat{\mathcal{X}}^{r}\right) = \prod_{(r, r', l, l') \in F^{r}} N(0, \Sigma_{\text{high}}) \doteq N(T_0, \Sigma_0). \tag{35}$$

Choosing the Most Probable Hypothesis

The decision on the most probable hypothesis is made as follows. The probabilities in (30) of only the most promising hypotheses are calculated, identified by their support (number of inliers). The highest probability hypothesis h is then chosen only if it satisfies two conditions: a) its posterior probability is sufficiently higher compared to all other hypotheses and b) its prior probability is significantly dominant with respect to other hypotheses.

Specifically, in the current implementation, a hypothesis is chosen only if the following two criteria that correspond to the above conditions are satisfied: a) the probability ratio between two most likely hypotheses [according to (30)] is above 2.0 (in other words, the most likely hypothesis $h^* \in \mathcal{H}$ is twice as probable than any other hypothesis) and b) the prior probability $p\left(h^* \middle| \hat{\mathcal{X}}^{r}\right)$ of the most likely hypothesis h^* is above 0.8. For numerical reasons, computation of $p\left(h \middle| \mathcal{Z}^{r}, \hat{\mathcal{X}}^{r}\right)$ is performed in the log space.

An interesting question is how much trajectory overlap, or mutually observed areas by different robots, is required for determining the initial relative pose. Unfortunately, there is no closed-form solution to this question since the answer depends both on the scenario and the tuning parameter α in the hypothesis prior. In particular, without perceptual aliasing, which is the case when there is low appearance similarity between different areas, there will be only a single cluster. On the other hand, in the presence of perceptual aliasing (for example, two corridors that look alike) there will be multiple clusters. In both cases, the CRP determines how much information must be accumulated before a decision is to be made regarding which cluster to use. The CRP is also used in the case when there is only a single cluster, since the cluster may represent consistent incorrect matches due to perceptual aliasing while the robots have not yet mutually observed any scene. In practice, robots typically have to travel through at least the same two corridors until a decision is made [see, for example, Figure 9(d)]. However, the actual trajectory overlap is scenario and tuning dependent (parameter α).

Incorporating Uncertainty Within Expectation Maximization

The EM formulation is extended here to incorporate uncertainties over robot poses. The derivation begins with the MAP inference in (37), which is now also conditioned on the covariance Σ^r

$$\hat{T}^r_{r'} = \arg\max_{T^r_{r'}} \sum_{\mathcal{J}^r} p\left(T^r_{r'}, \mathcal{J}^r \,\middle|\, \hat{\mathcal{X}}^r, \mathcal{Z}^r, \Sigma^r\right), \tag{S14}$$

and the corresponding EM formulation in (38)

$$\hat{T}^{(t)} = \arg\max_{T} \mathbb{E}_{\mathcal{J}^r \mid \hat{T}^{(t-1)}, \hat{\mathcal{X}}^r, \mathcal{Z}^r, \Sigma^r} \log\left[p\left(T, \mathcal{J}^r \,\middle|\, \hat{\mathcal{X}}^r, \mathcal{Z}^r, \Sigma^r\right)\right]. \tag{S15}$$

The covariance Σ^r represents the second moment of the pdf $p(\mathcal{X}^r \mid \mathcal{Z}^r)$ and can be evaluated as part of the MAP estimation in (3).

For brevity, the superscripts and subscripts r and r' are omitted from now on. Thus, the lower bound (S10) becomes

$$Q(T \mid \hat{T}^{(t)}) = \sum_{J} p\left(\mathcal{J} \,\middle|\, \hat{T}^{(t)}, \hat{\mathcal{X}}, \mathcal{Z}, \Sigma\right) \log p\left(T, \mathcal{J} \,\middle|\, \hat{\mathcal{X}}, \mathcal{Z}, \Sigma\right), \tag{S16}$$

and the explicit expressions of this bound, derived in "Expectation Maximization Equations," also become conditioned on Σ. In particular, the measurement likelihood changes from (S13) to

$$p\left(u^{r,r'}_{l,l'} \mid \hat{x}^r_l, \hat{x}^{r'}_{l'}, j_s, \hat{T}^{(t)}, \Sigma\right) \propto \exp\left(-\frac{1}{2} \| \operatorname{err}(u^{r,r'}_{l,l'}, \hat{x}^r_l, \hat{x}^{r'}_{l'}) \|^2_{\Sigma_{vs}}\right), \tag{S17}$$

which now includes a *modified* measurement covariance Σ_{vs} instead of the covariance $\Sigma_{js} \in \{\Sigma_{inlier}, \Sigma_{outlier}\}$ that appears in (S13).

The derivation proceeds by considering an inlier covariance Σ_{inlier}. The same procedure can be also applied for modifying the outlier covariance; however, since the latter is typically very large, it is typically not required in practice.

It is now necessary to develop an expression for the modified inlier covariance, Σ_{vs}. The predicted measurement u^{pred}_s, given the estimates $\hat{T}^{(t)}, \hat{x}^r_l, \hat{x}^{r'}_{l'}$ can be written as

$$u^{pred}_s = \hat{x}^r_l \ominus (\hat{T}^{(t)} \oplus \hat{x}^{r'}_{l'}) = h(\hat{T}^{(t)}, \hat{x}^r_l, \hat{x}^{r'}_{l'}), \tag{S18}$$

resulting in the measurement model

$$u_s = h(T, \hat{x}^r_l, \hat{x}^{r'}_{l'}) + v_s, \tag{S19}$$

with $v \sim N(0, \Sigma_{vs})$.

Furthermore, according to (S14), $T \equiv T^{r'}_r$ is the random variable in the considered inference problem, while the estimates $\hat{x}^r_l$ and $\hat{x}^{r'}_{l'}$ are given and fixed, and therefore *not* part of the inference.

At the same time, however, the inlier covariance Σ_{inlier} represents only the measurement noise v_s and not the errors in these estimates. In other words, this covariance models correctly the error $u_s - h(T^{r'}_r, x^r_l, x^{r'}_{l'})$ only for true values $\bar{x}^r_l$ and $\bar{x}^{r'}_{l'}$

$$u_s = h(T, \bar{x}^r_l, \bar{x}^{r'}_{l'}) + v_{inlier}, \tag{S20}$$

with $v_{inlier} \sim N(0, \Sigma_{inlier})$.

Therefore, the idea is to modify the covariance Σ_{inlier} (and similarly $\Sigma_{outlier}$) to account for the estimation errors in x^r_l and $x^{r'}_{l'}$. This modified covariance is Σ_{vs} in (S17); see a similar treatment in [23].

The terms Δx^r_l and $\Delta x^{r'}_{l'}$ may be introduced to account for errors in the estimates $\hat{x}^r_l$ and $\hat{x}^{r'}_{l'}$

$$\bar{x}^r_l = \hat{x}^r_l + \Delta x^r_l, \bar{x}^{r'}_{l'} = \hat{x}^{r'}_{l'} + \Delta x^{r'}_{l'}. \tag{S21}$$

Linearizing $h(T^{r'}_r, \bar{x}^r_l, \bar{x}^{r'}_{l'})$ from (S20) about $\hat{x}^r_l$ and $\hat{x}^{r'}_{l'}$ yields

$$u_s \approx h(T, \hat{x}^r_l, \hat{x}^{r'}_{l'}) + \frac{\partial h}{\partial x^r_l} \Delta x^r_l + \frac{\partial h}{\partial x^{r'}_{l'}} \Delta x^{r'}_{l'} + v_{inlier}. \tag{S22}$$

The uncertainty of the last three terms in (S22) is quantified by Σ_{vs} from (S17), which is therefore equal to

$$\Sigma_{vs} = \begin{pmatrix} \frac{\partial h}{\partial x^r_l} & \frac{\partial h}{\partial x^{r'}_{l'}} \end{pmatrix} \Sigma_{x^r_l, x^{r'}_{l'}} \begin{pmatrix} \frac{\partial h}{\partial x^r_l} \\ \frac{\partial h}{\partial x^{r'}_{l'}} \end{pmatrix} + \Sigma_{inlier}, \tag{S23}$$

where $\Sigma_{x^r_l, x^{r'}_{l'}}$ is the joint marginal covariance of x^r_l and $x^{r'}_{l'}$. This covariance can be extracted from the joint covariance $\Sigma \equiv \Sigma^r$. Calculating this joint covariance can be done efficiently while exploiting sparsity [S3].

REFERENCE

[S3] M. Kaess and F. Dellaert, "Covariance recovery from a square root information matrix for data association," *Robot. Auton. Syst.*, vol. 57, no. 12, p. 1198, 2009.

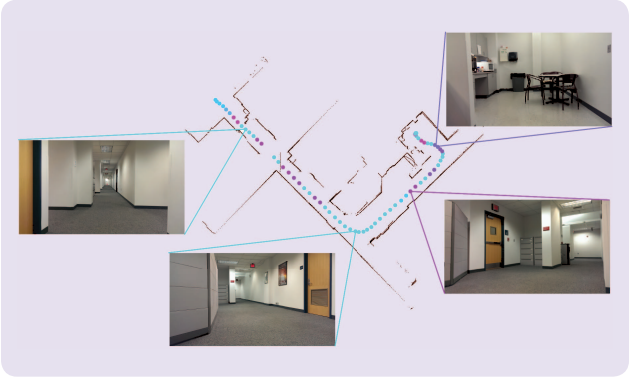


FIGURE 11 An autocovariance of laser scans collected along a robot's trajectory through a hallway environment. Magenta corresponds to laser scans with high autocovariance and therefore high saliency. Cyan corresponds to less salient laser scans.

TABLE 3 Worst-case time complexities for core algorithm components.

Component	Complexity	Phase
Correspondence generation	$O(n)/O(1)$	1, 2
Hypothesis generation	$O(\lvert\mathcal{F}^r\rvert)$	1
Hypothesis selection	$O(1)$	1
EM optimization	$O(\lvert\mathcal{F}^r\rvert)$	1, 2
Pose-graph optimization	$< O(n)$	1, 2

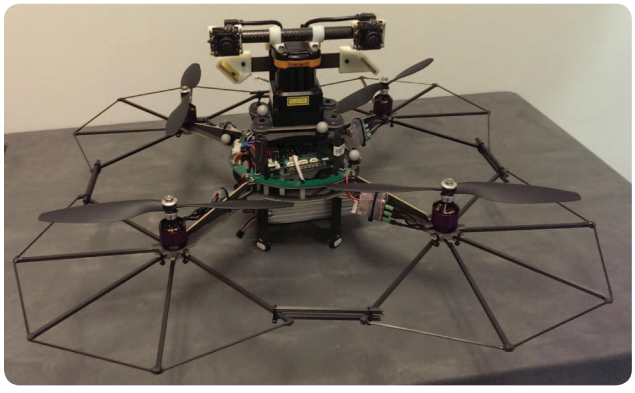


FIGURE 12 One of the aerial robots used for the experiments. The quadrotor platform is equipped with an onboard computer, inertial measurement unit, laser scanner, and beam deflector mirrors. Stereo cameras were not used in the described experiments.

INFERENCE OVER ROBOT TRAJECTORIES

After common reference frames between robot r and other robots r' are estimated using the approach described above, it becomes possible for robot r to infer its own and other robot poses with (3),

$$\hat{\mathcal{X}}^r = \arg\max_{\mathcal{X}^r} p(\mathcal{X}^r \mid \mathcal{Z}^r), \tag{36}$$

using multirobot constraints. Of course, at each point, this inference problem can be solved considering only those robots r' with whom robot r has established a common reference frame and excluding the rest. For clarity in this discussion, recall the original multirobot inference formulation from (8). The inferred common reference frames between robot r and other robots r' are used in the likelihood term of (8) [see (10)]. This inference problem can be solved with standard nonlinear optimization methods while using only the *inlier* multirobot data associations in $\mathcal{F}^r$, as identified by the EM approach. Alternatively, all the candidate constraints represented by the set $\mathcal{F}^r$ can be considered with the aim of solving the inference problem in (14). To that end, a similar EM approach can be developed; see [9]. In practice, the two approaches yield similar results.

Once a common reference frame is established between two robots, information sharing becomes easier. It becomes possible to match $z_k^{r'}$ against only a subset of sensor observations in $\{z_i^r\}$, skipping all observations z_i^r that were acquired too far away from x_k^r.

However, in both cases, due to drift in the robot pose estimates, identifying true multirobot inlier correspondences as inliers becomes very challenging over time, in particular when operating in large-scale environments. The next section presents a possible approach to overcome this challenge.

Multirobot Loop Closures in Large-Scale Environments

The problematic aspect in both the exact and EM formulation in (16) and (17) is that both are conditioned only on robot estimated poses, without accounting for the uncertainties in these estimations. This assumption is reasonable if estimation errors of each robot are small, such as when operating in small areas, or in scenarios where robots reduce estimation errors by performing loop-closure measurements. However, when operating in larger environments, robot pose estimations develop significant drift over time and cannot be considered negligible. In these scenarios, the presented EM formulation may not identify multirobot inlier correspondences even if those do exist.

An example is shown in Figure 10(a), where the blue and green robots start from same location, move toward the top left, diverge, and finally move into the same corridor where multirobot constraints can be formed. However, due to trajectory drift, the predicted relative pose is far from the measured relative pose (determined by true inlier multirobot correspondences). As a result, all of these inlier correspondences are identified as outliers, leading to suboptimal map merging and estimation accuracy.

One way to address this issue is to incorporate the uncertainty in robot poses within the EM formulation and accounting for this information when inferring if a candidate multirobot correspondence from the set $\mathcal{F}^r$ is an inlier or outlier. In other words, this provides a probabilistically sound mechanism to adapt the inlier and outlier

TABLE 4 The format of messages sent between robots. If the laser scan is not selected to be broadcast, only the first three items will be sent. *n* indicates the number of correspondences found for this laser scan.

Content	Size (bytes)
Robot ID	1
Pose index	4
Odometry	48
Laser scan	5800
Matched robot's ID	$n \times 1$
Matched pose's index	$n \times 4$
Matched relative pose	$n \times 48$
Total	$5853 + n \times 53$

covariances Σ_{inlier}, Σ_{outlier} according to the associated uncertainty in appropriate robot poses involved in each candidate correspondence in $\mathcal{F}^r$.

Specifically, the MAP estimation in (16) is now also conditioned on the joint covariance Σ^r available to robot r

$$\hat{T}_r^{r'} = \arg\max_{T_r^{r'}} \sum_{\mathcal{J}^r} p\left(T_r^{r'}, \mathcal{J}^r \middle| \hat{\mathcal{X}}^r, \mathcal{Z}^r, \Sigma^r\right), \tag{37}$$

and the EM formulation in (17) changes accordingly

$$\hat{T}^{(t)} = \arg\max_{T} \mathbb{E}_{\mathcal{J}^r \middle| \hat{T}^{(t-1)}, \hat{\mathcal{X}}^r, \mathcal{Z}^r, \Sigma^r} \log\left[p\left(T, \mathcal{J}^r \middle| \hat{\mathcal{X}}^r, \mathcal{Z}^r, \Sigma^r\right)\right], \tag{38}$$

where, like earlier, t denotes iteration number. The covariance Σ^r represents the second moment of the probability density function $p(\mathcal{X}^r | \mathcal{Z}^r)$ and can be obtained as part of the MAP inference in (3). The full derivation of incorporating uncertainty in robot poses within the EM framework is given in "Incorporating Uncertainty Within Expectation Maximization." Such an approach allows us to overcome the above described challenges and correctly identify multirobot inlier correspondences even when the estimated robot poses develop significant drift.

COMPUTATIONAL CONSIDERATIONS

This section describes an information-based method for downsampling the number of sensor observations shared among robots, as well as a computational complexity analysis of the formulation.

Observation Subsampling and Saliency

To estimate the initial relative poses of all robots, the data-association strategy requires that each robot compares sensor observations shared by other robots with its own local history of observations. Although the comparison itself can be performed efficiently through ICP scan matching, feature-based matching, or other robust correspondence techniques, the number of comparisons that must be performed for each update grows linearly in the number of locally stored sensor observations. Unfortunately, it is not possible to employ strategies such as comparing against only the previous laser scan captured by each robot since robots may observe the same location in the environment at different times.

TABLE 5 Data sets used to evaluate the approach.

Data Set	Type	Robots	Size
D1	Indoor	3	15 m × 10 m
D2	Indoor	3	30 m × 20 m
D3	Indoor	3	30 m × 30 m
D4	Indoor	3	70 m × 50 m
D5	Indoor + outdoor	2	60 m × 50 m
D6	Outdoor	3	30 m × 30 m

Although the linearly increasing number of range observations comparisons is necessary to correctly infer the relative transform between robots, each robot does not need to share all of its collected sensor observations with other robots. Instead, observations can be subsampled to reduce both the computational and network cost. While a naïve subsampling approach might uniformly select laser scans to share with other robots, a more efficient approach would only share a subset of collected laser scans that are most salient and, therefore, most descriptive of the environment from which they were captured. The autocovariance of a laser scan is calculated [44] to determine saliency. Autocovariance is calculated per laser scan, prior to sharing with other robots. To calculate laser scan autocovariance, each new laser scan is randomly perturbed with Gaussian noise and matched with itself several times to produce a transformation estimate τ_i with covariance Σ_i. After N such iterations, saliency is computed as $(\text{tr}(\Sigma))^{-1}$ according to

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(\Sigma_i + (\tau_i - \mu)(\tau_i - \mu)^T), \qquad \mu = \frac{1}{N}\sum_{i=1}^{N}\tau_i. \tag{39}$$

For additional details, the reader is referred to a similar article using autocovariance for observation saliency [44]. Figure 11 illustrates laser scan autocovariance, calculated using the above approach, for several locations in a self-similar hallway environment.

Timing Complexity of the Approach

The multirobot data association and distributed-inference strategy includes two main phases of operation with differing algorithmic complexities. Although all components of the algorithm are roughly linear in the number of stored laser scans or poses in each robot's trajectory, the two

phases involve different components (Table 3). During the first phase, transformations between robot coordinate frames are established by determining correspondences between each robot's history of observations and incoming observations shared by other robots. This phase consists of four components: correspondence generation, hypothesis generation and selection, EM, and pose-graph optimization. During the second phase, transformations have already been established, and each robot's pose graph is optimized using new measurements and correspondences. The second phase still requires correspondence generation, EM, and pose-graph optimization but no longer requires hypothesis generation and selection.

Correspondence generation involves comparing incoming laser scans that are shared by other robots with a local history of stored sensor observations. Before the common reference frame is built between two robots, if a robot locally stores one laser scan per pose estimate, then this step has $O(n)$ complexity, where n is the number of poses in the robot's trajectory. After the common reference frame is built, since the robot already knows roughly the position of another robot, it only needs to match nearby scans from the other robot(s), and thus the complexity is reduced to $O(1)$.

Hypotheses are being generated until the prior of at least one hypothesis is large enough to make a decision regarding the transformation between the two robots. This

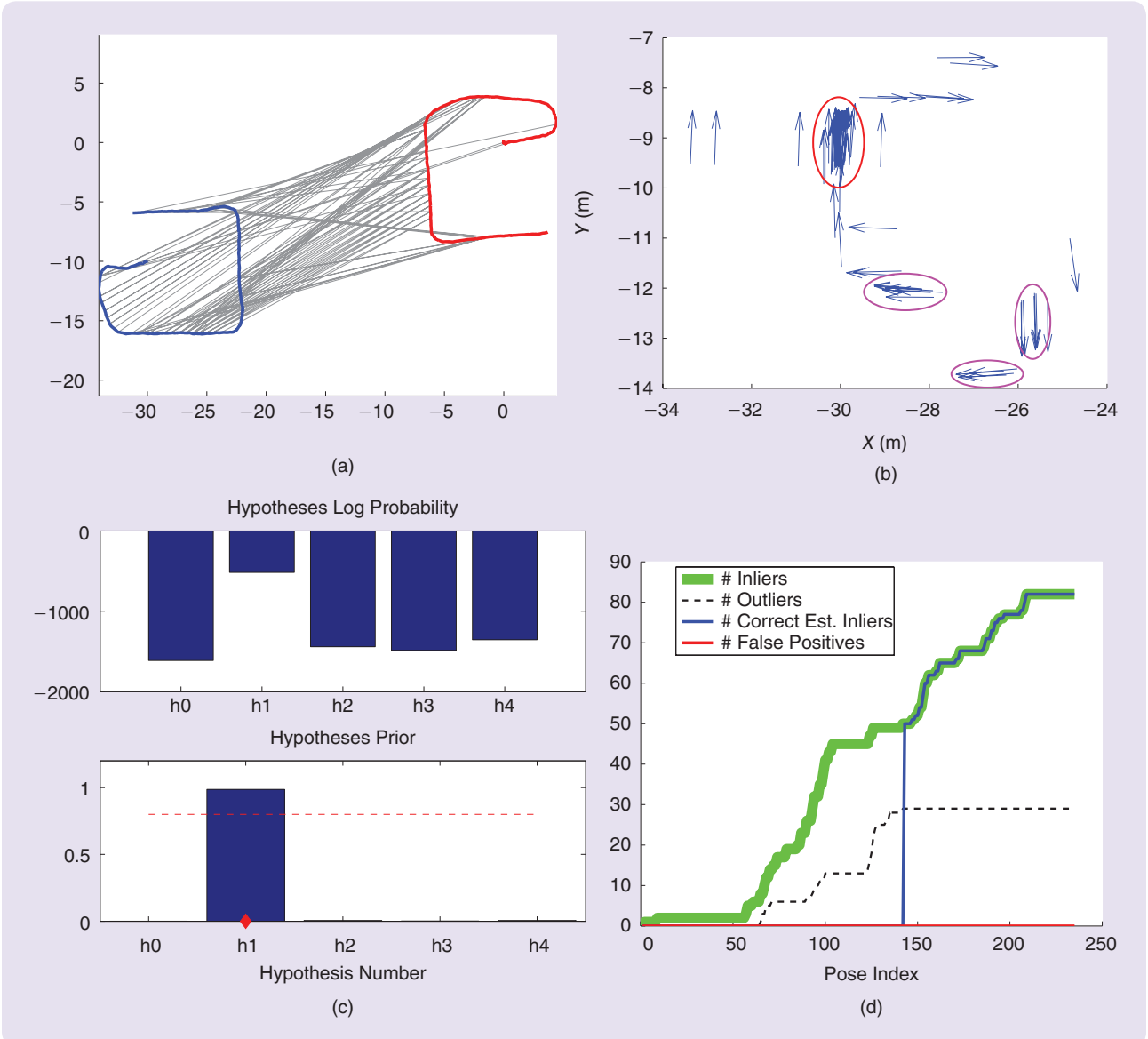


FIGURE 13 Data set **D1**: (a) The candidate correspondences between the red and blue robots. Initial poses of these robots are set to arbitrary values. (b) The distribution of the set $T_r^r(\mathcal{F}^r)$ for red and blue robots. The correct hypothesis is marked by a red circle, and the three incorrect hypotheses are marked in magenta circles. (c) Hypothesis probability (in log-space) and prior. The hypothesis h_1, corresponding to the dominant cluster (marked by a red circle) in (b), is chosen. h_0 represents the null hypothesis so it does not have a corresponding cluster in (b). (d) The actual and inferred inliers and outliers.

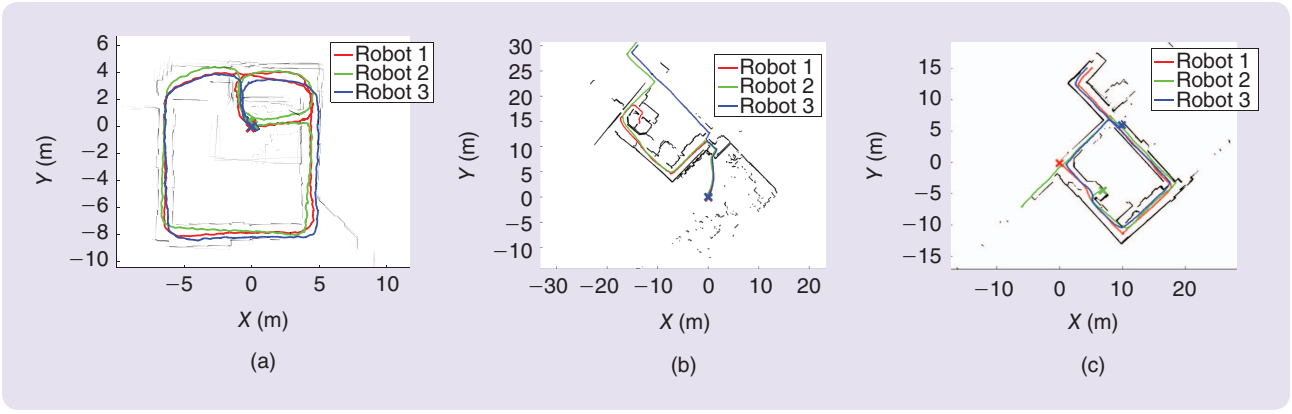


FIGURE 14 Ground truth for the data sets (a) **D1**, (b) **D2**, and (c) **D3**. The robot's initial positions are denoted by cross marks.

step involves an EM optimization, which has a complexity of $O(|\mathcal{F}^r|)$, as detailed below. Generally, the set of hypotheses is small, containing between one and ten elements. Selecting the best hypothesis then involves iterating over the small set once, performing the $O(1)$ computation, and finding the maximum.

EM is performed after new correspondences are generated between robots. Since EM involves evaluating all correspondences in the multirobot pose graph, the complexity of each EM update is $O(|\mathcal{F}^r|)$. While this step has a linear complexity, the number of correspondences in each robot's pose graph is guaranteed to be lower than the number of poses in the graph ($n > |\mathcal{F}^r|$). In addition, EM is only performed once a new correspondence has been generated, making it infrequent in comparison to correspondence generation.

Finally, pose-graph optimization involves iterating over any poses affected by new multirobot correspondences. A worst-case scenario, in which a dense pose graph contains many multirobot correspondences, has a $O(n)$ computational cost. However, in general, pose-graph optimization only involves incremental updates to several nodes [45] and is therefore be more efficient than $O(n)$.

PERFORMANCE STUDY

The formulation described above was implemented within the GTSAM library [46] to analyze accuracy and efficiency. The implementation was evaluated in real-time multirobot real-world experiments in both indoor and outdoor environments. This section describes implementation details, discusses the experimental setup, and analyzes results of the multirobot experiments.

Experimental Setup

Quadrotor aerial vehicles were used as robotic platforms for real-time experimentation (Figure 12). The quadrotor platforms were equipped with onboard computers (1.86-GHz Intel Core 2 Duo processors); Hokuyo URG-30LX 30-m range laser scanners; inertial measurement units; and downward-facing, laser-beam-deflector mirrors. Ground truth poses and maps were estimated using a three-dimensional SLAM framework [47], which leverages ICP for laser-based odometry, a histogram filter for localization, and an unscented Kalman filter for state estimation.

The approach was implemented in a distributed incremental framework. During real-time operations, each robot captured laser scans. All robots ran individual instances of the implementation to generate pose estimates in their local reference frames. Laser scans were User Datagram Protocol broadcast to all other robots on the wireless network in data packets containing the laser scan observation, a corresponding local pose estimate, and the broadcasting robot's identity.

Each robot r executed the same algorithm: each robot shared highly informative laser scans, received scans from other robots, and calculated relative pose constraints between these scans and its own informative scans using ICP.

TABLE 6 The initial relative-pose estimation errors in the three real-world data sets D1, D2, and D3. Estimation errors are reported in terms of norm of translation error ($\|t\|$) in meters and absolute value of orientation error ($|\theta|$) in degrees.

Estimation Error:			$\|t\|$	$\|\theta\|$	$\|t\|$	$\|\theta\|$	$\|t\|$	$\|\theta\|$
Data Set	Size	r, r'	Red		Green		Blue	
D1	15 × 15	Red	—	—	0.5	6.0	0.4	4.7
		Green	0.4	6.5	—	—	0.4	10.9
		Blue	0.4	4.9	0.3	10.4	—	—
D2	20 × 30	Red	—	—	0.1	1.4	0.3	1.8
		Green	0.1	1.2	—	—	0.4	3.2
		Blue	0.35	2.3	0.4	3.5	—	—
D3	30 × 30	Red	—	—	0.6	0.6	2.1	3.3
		Green	0.5	0.3	—	—	1.3	8.2
		Blue	2.7	3.4	2.9	8.6	—	—

A decision must eventually be made regarding whether sufficient information has been accumulated to choose a hypothesis.

The saliency of a scan was calculated according to its autocovariance. Incoming observations from other robots were matched against locally stored measurements, which were captured nearby to generate correspondences $\mathcal{F}^r$. Odometry was obtained by updating robot poses using incremental ICP transformations between local consecutive laser scans. Correspondences and odometry information was stored in the robot's local pose graph and optimized using GTSAM. The optimization ran at 20 Hz on each robot. Local odometry and robot pose estimates were broadcast over the network at 20 Hz and laser scans at up to 5 Hz. Table 4 gives the format of messages shared between robot. The size of each message is roughly 53 B with no laser scan and 6 kB if a laser scan is included.

In all cases, since the initial relative poses between the robots were unknown, the initial pose of each robot was set to an arbitrary value [see Figure 13(a)]. Furthermore, in all cases the parameter α was set to 500. Unless otherwise specified, robot local measurements do not include loop-closure constraints.

Data Sets

The approach was first evaluated using three real-world indoor data sets, **D1**, **D2**, and **D3**, captured by three quadrotors equipped with laser scanners. Reference trajectories, color coded according to robot number (red, green, blue), are shown in Figure 14 along with the laser scans of the first robot (red). In two of the data sets, **D1** and **D2**, the robots start operating from the same location. In the third data set, **D3,** the robots start operating from different locations. Additionally, the implementation was examined using three larger-scale data sets **D4**, **D5**, and **D6**, which include an indoor environment, an environment with a transition between indoors

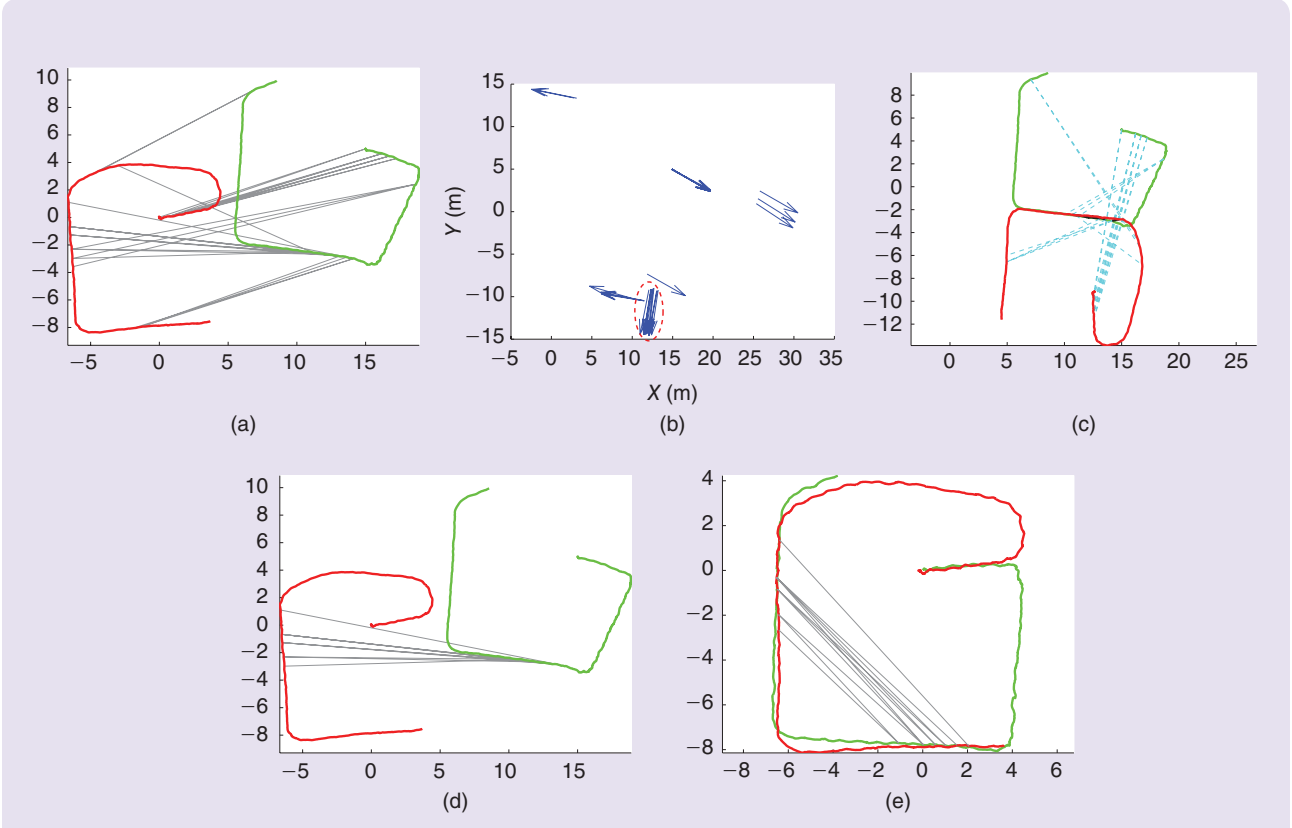


FIGURE 15 Data set **D1**: The effect of *measurement aliasing* when *not* using the hypothesis prior. (a)–(b) Candidate correspondences $\mathcal{F}^r$ and the distribution of set $T_r^{r'}(\mathcal{F}^r)$ for green and red robots $(r = 2, r' = 1)$. A cluster of consistent outliers is emphasized. (c) The expectation maximization optimization result using an incorrect hypothesis. Robot trajectories are erroneously aligned. (d) Inliers of the chosen hypothesis (all outliers in practice) drawn on top of robot trajectories with arbitrary initial pose and (e) on top of ground-truth trajectories.

and outdoors, and an exclusively outdoor environment. Table 5 summarizes the characteristics of the different data sets that were used in the evaluation.

Evaluation Metrics

This evaluation uses the following metrics:

- Estimation accuracy of the initial relative poses $T_{r'}^{r}$ between any two robots r and r'.
- Statistics of inferred multirobot data association. The percentage of correctly identified inliers and outliers, as well as false negatives (inliers that were identified as outliers) and false positives (outliers that were identified as inliers). Ground-truth inliers were determined by evaluating the error in each multirobot correspondence using ground-truth data and identifying errors smaller than a threshold.
- Robustness to measurement aliasing. Incorporating the CRP within the hypothesis prior allows the algorithm to reliably choose the correct hypothesis in an incremental framework, even in presence of measurement aliasing.
- Exhibit the ability of each robot to infer its own and other robots' trajectories, once appropriate initial relative poses $T_{r'}^{r}$ are established.

Results

A detailed performance analysis is provided for the first data set **D1** and results from the other data sets are also discussed.

Data Set **D1**

In this data set, the robots begin moving from the same position, with the red and blue robots moving counterclockwise, and the green robot moving in the clockwise direction [see the ground-truth trajectories in Figure 14(a)]. In contrast to the other data sets, this data set contained manually identified loop closures for each of the three robots.

Figure 13 details the algorithm steps considering establishing initial relative-pose transformations $T_{r'}^{r}$ between the red and blue robots (r = red and r' = blue). Figure 13(a) shows the candidate correspondences between the two robots (the set $\mathcal{F}^{r}$) that are available to the red robot just before the transformation $T_{r'}^{r}$ is estimated.

Recall that the set $\mathcal{T}_{r'}^{r}(\mathcal{F}^{r})$ represents the initial relative poses that are calculated for each of the candidate correspondences in $\mathcal{F}^{r}$. Figure 13(b) shows the distribution of this set $\mathcal{T}_{r'}^{r}(\mathcal{F}^{r})$, with each arrow representing a single planar transformation $T_{r'}^{r} \in \mathcal{T}_{r'}^{r}(\mathcal{F}^{r})$ in terms of (x, y) coordinates, where orientation θ is described by the arrow's angle.

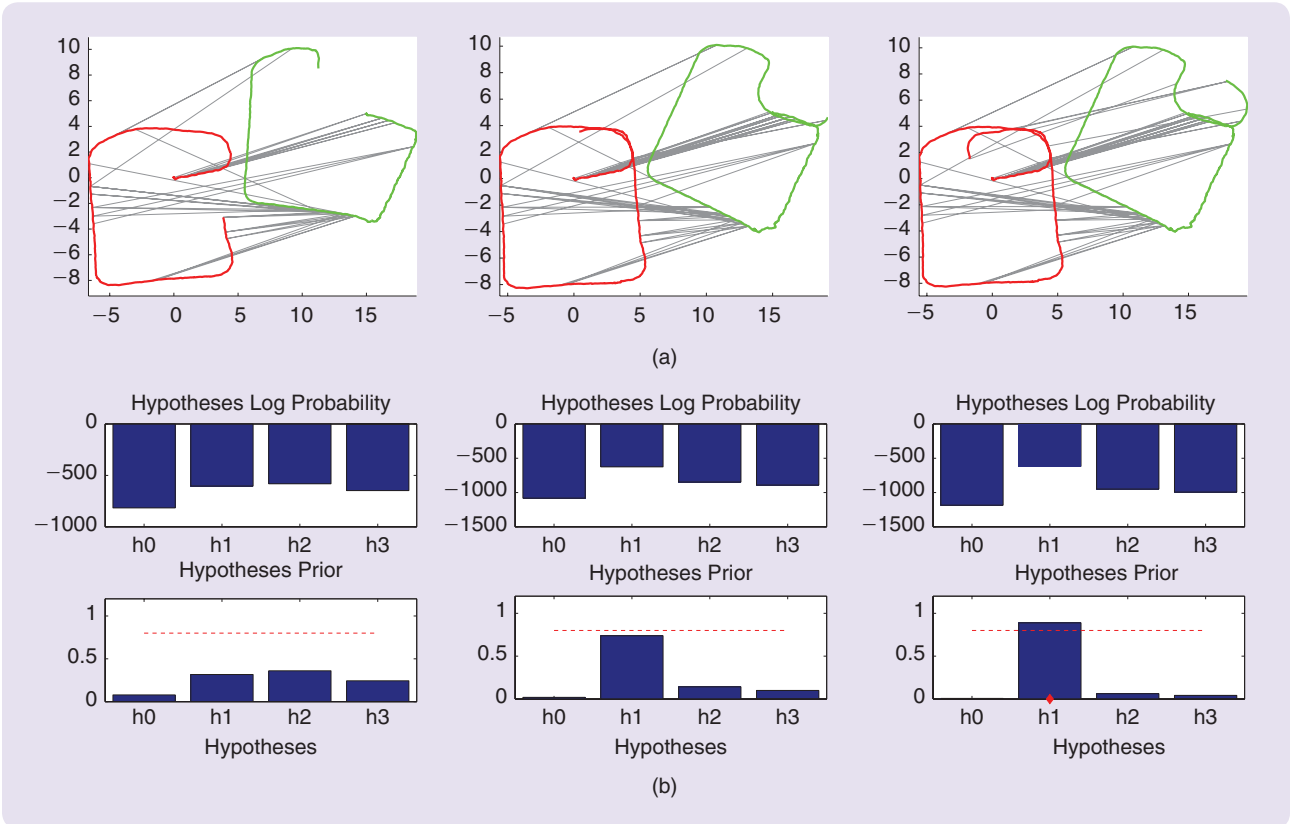


FIGURE 16 Data set **D1**: Establishing a common reference frame between green and red robots ($r = 2, r' = 1$). (a) Candidate correspondences $\mathcal{F}^{r}$ between green and red robots at three different time instances: indices 4778, 5690, and 6081. (b) Hypotheses posterior (top row) and prior (bottom row) probabilities for the three time indices. The posterior probability is shown in log space. The first hypothesis represents the null hypothesis. At time index 6081, the second hypothesis is selected because it has the highest posterior probability and its prior probability crosses the threshold (0.8).

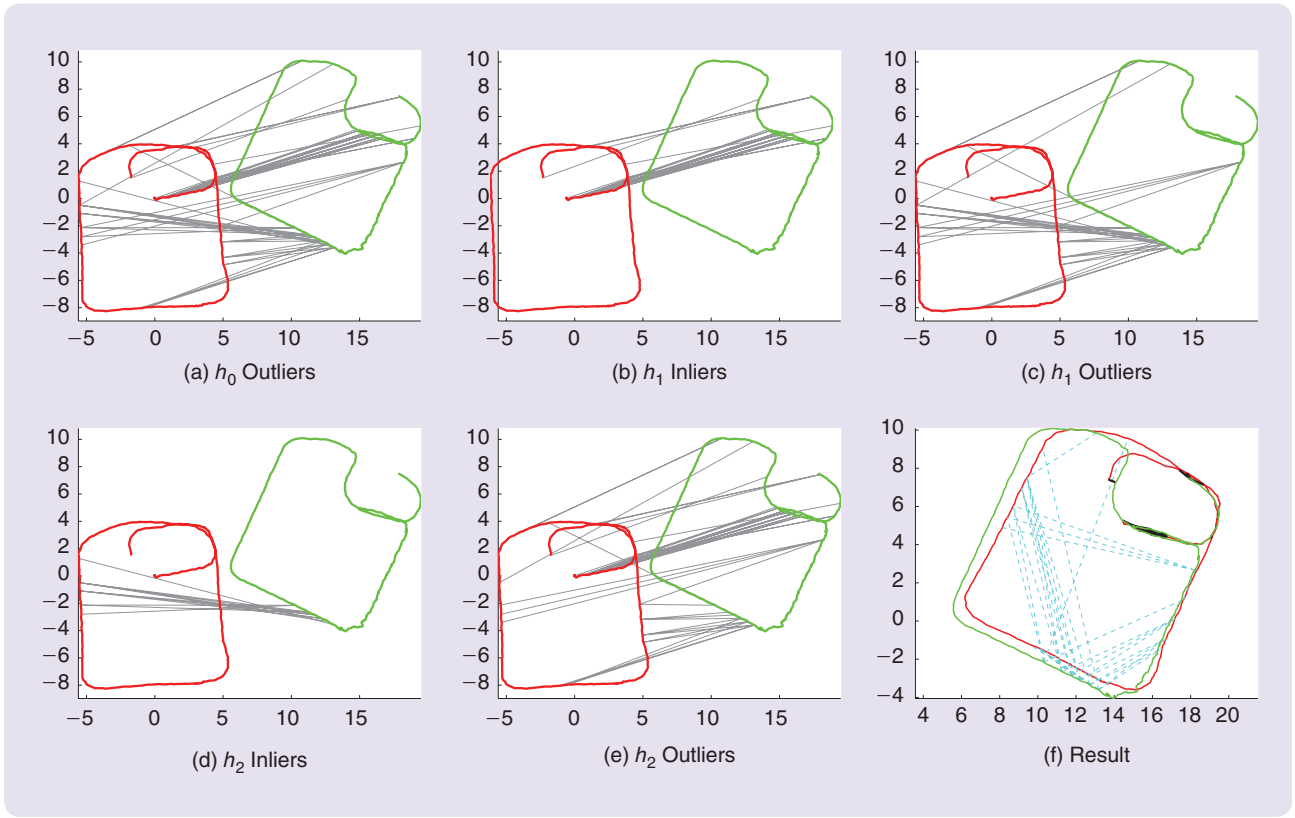


FIGURE 17 Data set **D1**: Details of prior probability calculation at time index 6081. (a) Outlier correspondences of the null hypothesis h_0 (no inliers by definition). (b) and (c) Inlier and outlier correspondences of the chosen hypothesis (h_1). (d) and (e) Inlier and outlier correspondences of another hypothesis (h_2). (f) Expectation maximization optimization result using the correctly chosen hypothesis [the ground truth is given in Figure 14(a)]. Inliers are denoted by the black color.

The probabilities of the identified hypotheses that correspond to the most dominant clusters in Figure 13(b) are shown in Figure 13(c). The probability of the null hypothesis h_0 is shown as well. Figure 13 shows both the posterior probability $p\left(h|\hat{\mathcal{X}}^r, \mathcal{Z}^r\right)$ (upper part) and the prior probability $p(h|\hat{\mathcal{X}}^r)$ (bottom part); see (30) and (33). The posterior probability is given in log-space and without the constant c. The threshold on the hypothesis prior for selecting the most probable hypothesis (0.8) is shown with a red dashed line. In this case, there are five hypotheses in the set $\mathcal{H}$. Since the posterior for h_1 is higher that the other hypotheses in $\mathcal{H}$ and its prior is above the threshold, this hypothesis is selected; $h^* \doteq h_1$.

Once a hypothesis is chosen, it becomes possible for robot r (the red robot) to perform the multirobot optimization in (36) based on all the multirobot constraints in $\mathcal{F}^r$ that involve robot r' (the blue robot) and possibly other robots with an established common frame. The r'th robot's poses in $\hat{\mathcal{X}}^r$ in (7) can now be expressed in the reference frame of robot r using $\hat{T}_r^{r'}(h^*)$. A correspondence $(r', k, l) \in \mathcal{F}^r$ is identified as inlier if $p\left(j_{k,l}^{r',r}|\hat{\mathcal{X}}^r, \mathcal{Z}^r\right)$ approaches 1.0. Figure 13(d) depicts a case in which the described approach correctly identified all inliers; in general, the method tends to infer a majority of the inliers. Furthermore, the method did not produce any false positive decisions (identified inliers that are outliers in practice).

The result of the multirobot pose-graph optimization in (36) is shown in Figure 9(a), with identified inliers indicated in black and outliers in gray. Robot trajectories are properly aligned with each other, compared to the reference trajectories in Figure 14(a). Estimation errors in the common reference frame between the two robots are given in Table 6.

After computing a common reference frame, new multirobot correspondences were directly added to the pose graph and optimized using (8). Results of this optimization for several time instances are shown in Figure 9(a) and(c) (only the identified inliers are shown in the last two figures). Similar results are obtained by the blue robot ($r = 3$) for $r' = 1$, as shown in Figure 9(d) and (f).

Establishing an initial relative pose and multirobot data association for the green robot is more challenging. It travels in the opposite direction (clockwise) with respect to the other two robots and therefore does not observe the same areas with these robots until the very end [the top right area in Figure 14(a)]. Moreover, since the robots operate in similar environments (for example, two corridors), measurement aliasing causes laser scans from these different environments to be correctly matched, suggesting the robots are

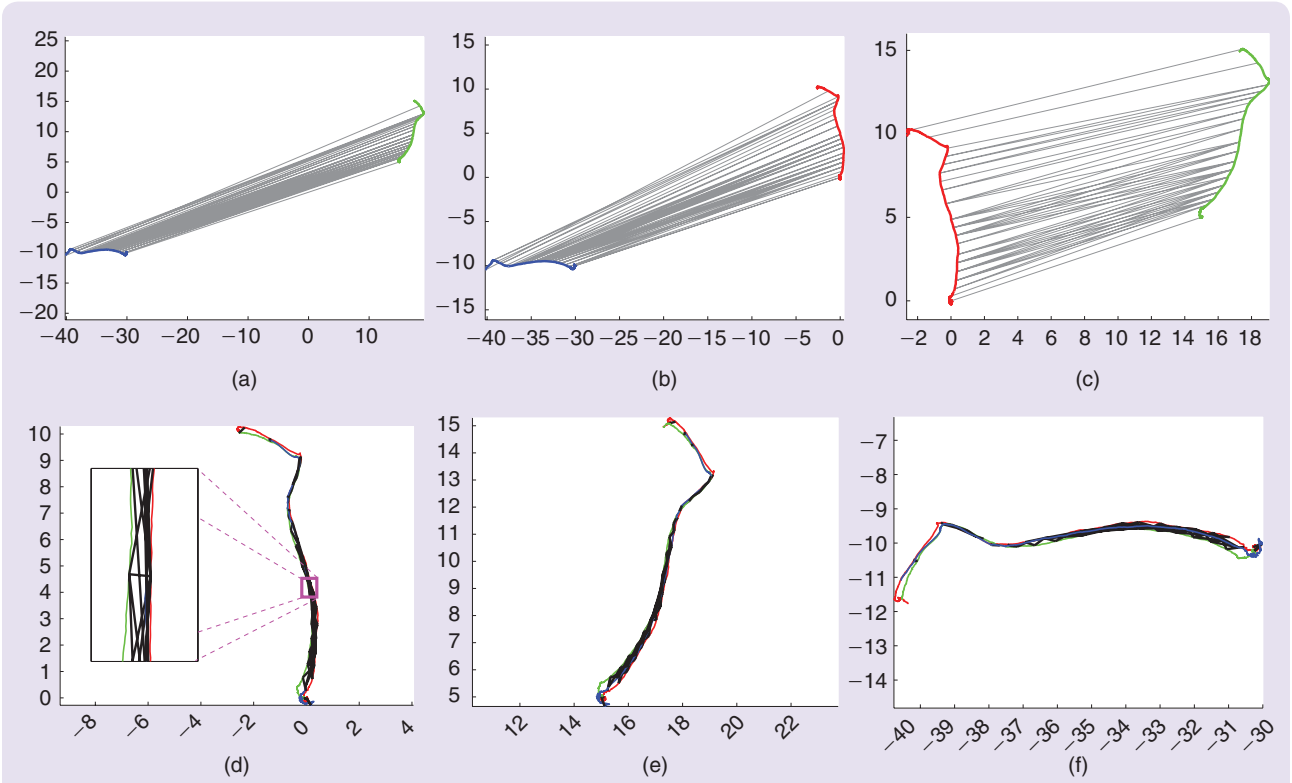


FIGURE 18 Data set **D2**. (a)–(c) Robot trajectories with an initial pose set to arbitrary values and multirobot candidate correspondences. (d)–(f) Identified inlier correspondences and robot trajectories expressed in the same reference frame once the latter is established. (a) Blue and green robots, (b) blue and red robots, (c) red and green robots, (d) the red robot (a zoom in on inliers is shown), (e) the green robot, and (f) the blue robot.

actually observing the same environment. As a result, a cluster of consistent outliers is obtained, and only after the robots actually observe the same environments is the cluster corresponding to the correct transformation formed.

Without using a hypothesis prior $p(h|\hat{\mathcal{X}}^r)$, there is nothing to prevent choosing the consistent-outlier hypothesis, which can lead to catastrophic results, as shown in Figure 15. Figure 15(a) and (b) shows, respectively, the candidate correspondences $\mathcal{F}^r$ between the green and red robots and the equivalent distribution of the initial relative poses (the set $\mathcal{T}_r^r(\mathcal{F}^r)$). The cluster corresponding to consistent outliers is indicated by a red ellipse. The inliers I of this consistent-outlier hypothesis $h = \{I, O\}$ are shown in Figure 15(d) on top of the robot trajectories expressed with arbitrary initial relative pose (since it is unknown); the same inliers are shown in Figure 15(e) on top of the ground-truth trajectories. All of these correspondences are erroneously considered as inliers since they relate between different areas. Using the chosen (incorrect) hypothesis and optimizing with (36) produces the result given in Figure 15(c), where the trajectories of the green and red robots are incorrectly aligned.

In contrast, incorporating the hypothesis prior $p(h|\hat{\mathcal{X}}^r)$ prevents making this incorrect decision. By introducing the null hypothesis and modeling the probability of observed unique places for each hypothesis, the priors for different hypotheses (including consistent-outlier and null hypotheses) compete with each other, and only after accumulating more observations by covering additional areas, the prior corresponding to a hypothesis with the highest number of inliers exceeds the threshold and therefore is selected. However, by that time the robots have already observed common areas, and the chosen hypothesis is indeed the correct one.

TABLE 7 Data set D1: The number of involved unique places for the hypotheses shown in Figure 17, and the corresponding *unnormalized* prior $f(n)$ from (S8). The normalized prior probabilities are shown in the last column as well as in Figure 16.

h	m	m_{in}	m_{out}	n	$f(n)$	Prior
h_0	58	0	58	116	4e-6	0.007
h_1	58	26	32	90	5e-5	0.89
h_2	58	11	47	105	3e-5	0.06
h_3	58	9	49	107	2e-5	0.04

The problem of choosing the most probable solution is particularly important in the incremental setting in the presence of measurement aliasing.

Deciding when to consider if sufficient information has been accumulated is determined by a hypothesis prior exceeding a threshold. This decision depends on the value of the parameter α, since it determines, given the competing hypotheses in $\mathcal{H}$, how fast the prior of a hypothesis becomes dominant. α must be set according to some problem-specific knowledge, such as expected size of the environment to be traveled. In practice, it is better to use conservative (large) values of α to reduce chances of making the incorrect decision.

The time evolution of the hypothesis priors of the green and red robots is illustrated in Figures 15 and 16 ($\alpha = 500$ in all cases). Three different time instances, represented by indices 4778, 5690, and 6081, are considered. Figure 16(a) shows the candidate correspondences in these time instances on top of the evolving robot trajectories, and (b) details posterior (in log space) and prior probabilities of the identified hypotheses. The hypothesis h_1 has the highest posterior probability starting from time index 5690, but its prior reaches the threshold only at time index 6081. Therefore, only at that time this hypothesis is chosen.

Figure 17 provides further details for the underlying process at time index 6081. The inlier and outlier correspondences for the null hypothesis h_0, the chosen hypothesis h_1, and another hypothesis h_2 are shown in Figure 16(a)–(e). Note that only outliers are shown for the null hypothesis, since by definition it does not contain any inliers. Table 7 further details the different parameters that are used for calculating the prior probability: number of inliers m_{in} and outliers m_{out}, total number of correspondences $m = m_{\text{in}} + m_{\text{out}}$, and the number of unique places n. Finally, the result of the EM optimization using the correctly chosen hypothesis is shown in Figure 17(f).

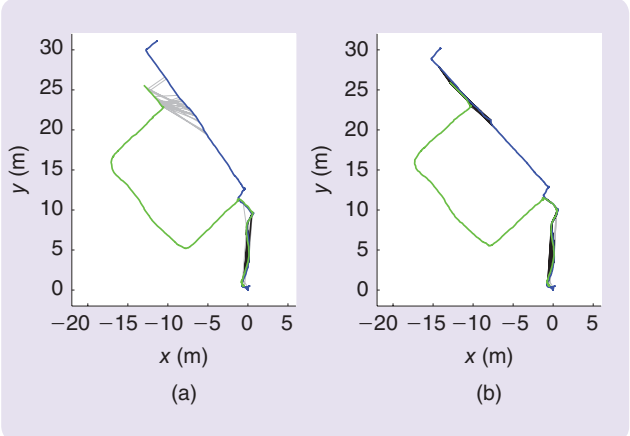


FIGURE 19 The green and blue robot trajectories (a) before and (b) after compensating for odometry measurements uncertainties in the expectation maximization optimization. The black lines signify inlier correspondences, while gray lines signify outliers. Before considering odometric uncertainty, correspondences are incorrectly classified as outliers due to significant odometric drift. After considering odometric uncertainty, inlier correspondences are correctly classified. (a) Without a noise-model update and (b) with a noise-model update.

Data Set **D2**

Figures 18 and 10 describe the results for data set **D2**. Figure 18(a) shows the candidate correspondences (the sets $\mathcal{F}^r$) between different pairs of robots. In this particular case, there was only one hypothesis (excluding the null hypothesis) that indicated that all of these correspondences were inliers. The prior of this hypothesis and the null hypothesis competed with each other, and after a sufficient number of observations the prior of the former reached the predefined threshold (0.8) and therefore this hypothesis was chosen. Similar to the basic example from Figure 8, for a given m (number of correspondences) at each time step, the number of unique places for the chosen and null-space hypotheses are, respectively, m and $2m$.

Figure 18(b) shows the result of the EM optimization in (36), for each of the three robots, right after choosing the hypothesis as described above. The algorithm correctly determined that the robots begin from the same location. Estimation errors of initial relative poses between the robots are provided in Table 6.

The estimates from all three robots are shown in Figure 10. Since no loop-closure constraints were used in this experiment, drift is expected in the trajectories of individual robots. Figure 10(a) shows the results from three robots without considering pose uncertainty. Compared to ground truth in Figure 11(b), the output trajectories and map are affected by odometry drift. Even though multirobot correspondences are explored between blue and green robots, they are treated as outliers and not considered in the inference. In Figure 19(a) the correspondences on the top (gray) are identified as outliers. In Figure 19(b) the inference includes robot pose uncertainties, and the same correspondences are instead identified as inliers. The effects of introducing robot pose uncertainty into the joint inference are shown for all three robots in Figure 10(b).

Data Set **D3**

Algorithm performance in the third data set **D3** is summarized in Figures 20 and 21 and in Table 6. In contrast to

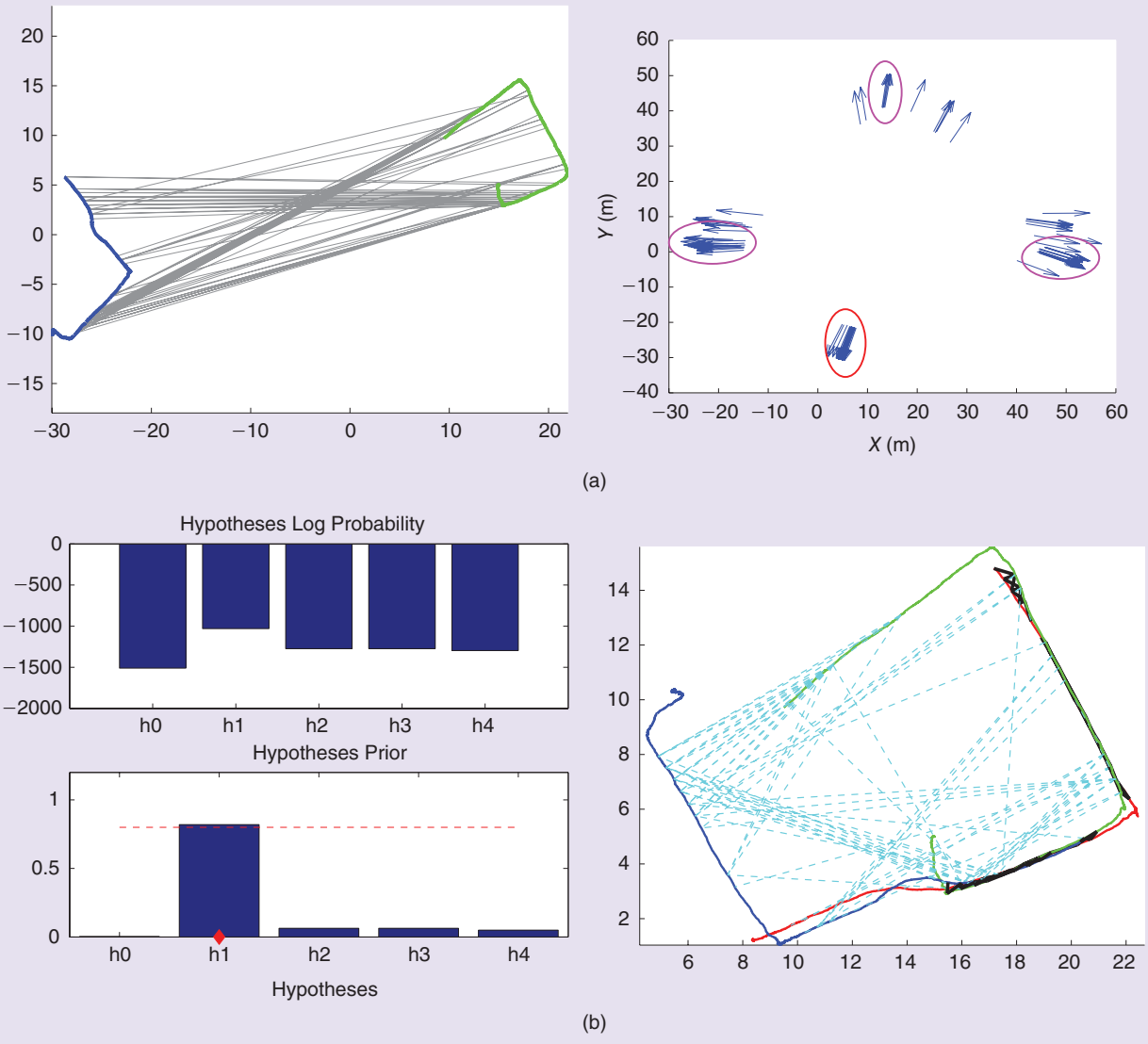


FIGURE 20 Data set **D3**: (a) Candidate correspondences and the equivalent distribution of the initial relative poses between green and blue robots. The correct hypothesis is marked by a red circle, and the three wrong hypothesis are marked in magenta circles. (b) Hypotheses (posterior and prior) probabilities and the result of multirobot optimization in (36) after hypothesis h_1 is chosen.

the first two data sets, the robots in **D3** begin operating from different positions [see Figure 14(c)]. The environment contains structural symmetries, such as several long corridors and similar corners. This problem is challenging, since perceptual aliasing gives rise to consistent outliers, as previously shown in Figure 3. Nevertheless, the described approach is capable of identifying the correct initial relative transformation between the robots. Figure 20(a) shows the set of candidate correspondences $\mathcal{F}^r$ and the equivalent distribution of the initial relative poses (the set $\mathcal{T}_r^{r'}(\mathcal{F}^r)$) between the green and the blue robots. The latter exhibits several clusters, and the most dominant four clusters were identified. The probabilities of the corresponding hypotheses including the null hypothesis are shown in Figure 20(b), and the hypothesis h_1 is chosen. The result of the EM multirobot optimization is shown in the right side of Figure 20(b) (the trajectory of the red robot is also shown since, by that time, a common reference frame between the green and the red robot has been established). The robot trajectories expressed in the green robot's reference frame are aligned well with the ground-truth trajectories [Figure 14(c)].

Similar to data set **D2**, since no loop closures were used, the trajectories drift over time if pose uncertainties are not considered [Figure 21(a)]. The drift is corrected for after considering pose uncertainties [Figure 21(b)].

Large-Scale Data Sets

To further evaluate the robustness of the multirobot inference approach, three more experiments in larger environments

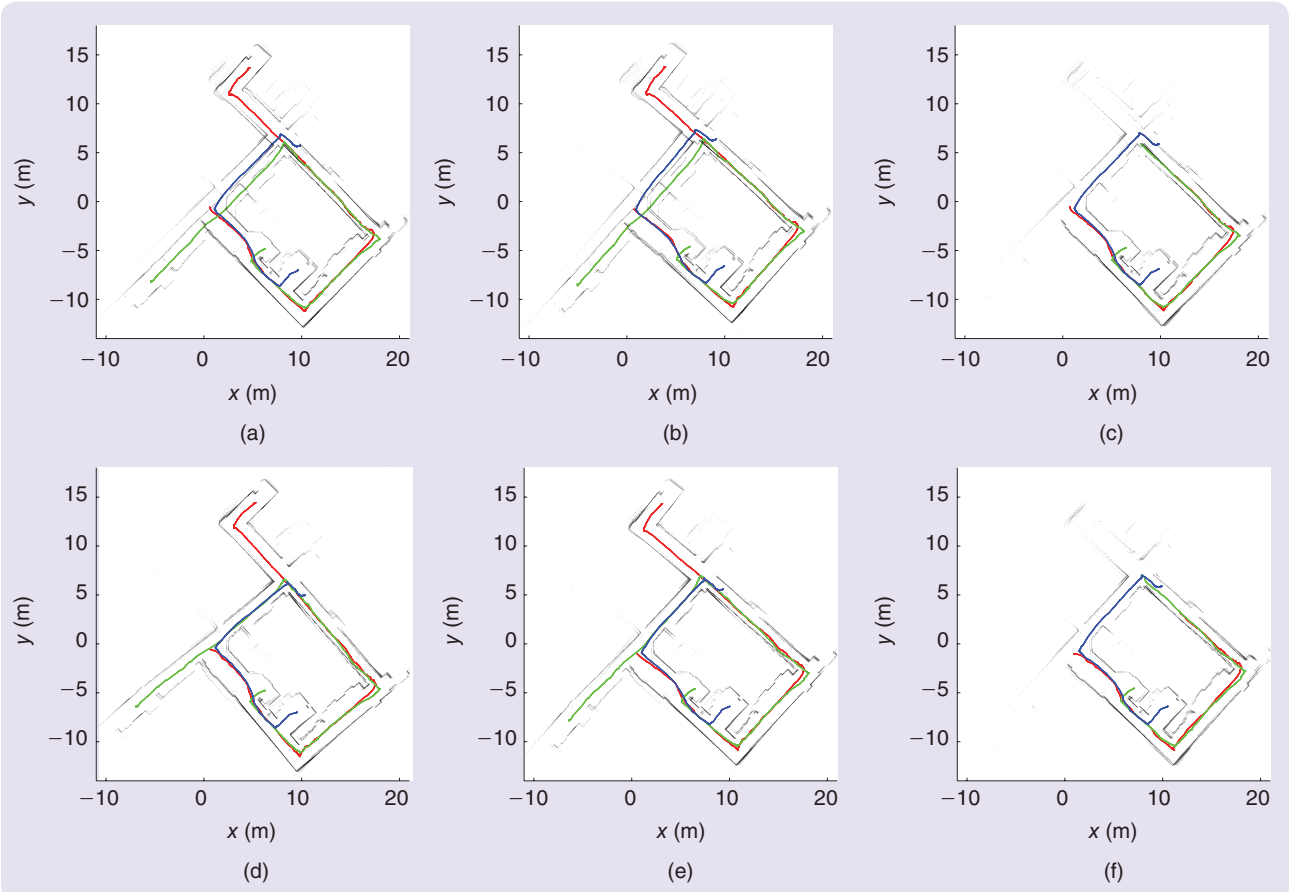


FIGURE 21 Data set **D3**: (a)–(c) The final estimated trajectories of all three robots, output by each robots, without considering odometric uncertainty; (d)–(f) estimated trajectories with considering odometric uncertainty. (a) By the red robot, (b) by the green robot, (c) by the blue robot, (d) by the red robot, (e) by the green robot, and (f) by the blue robot.

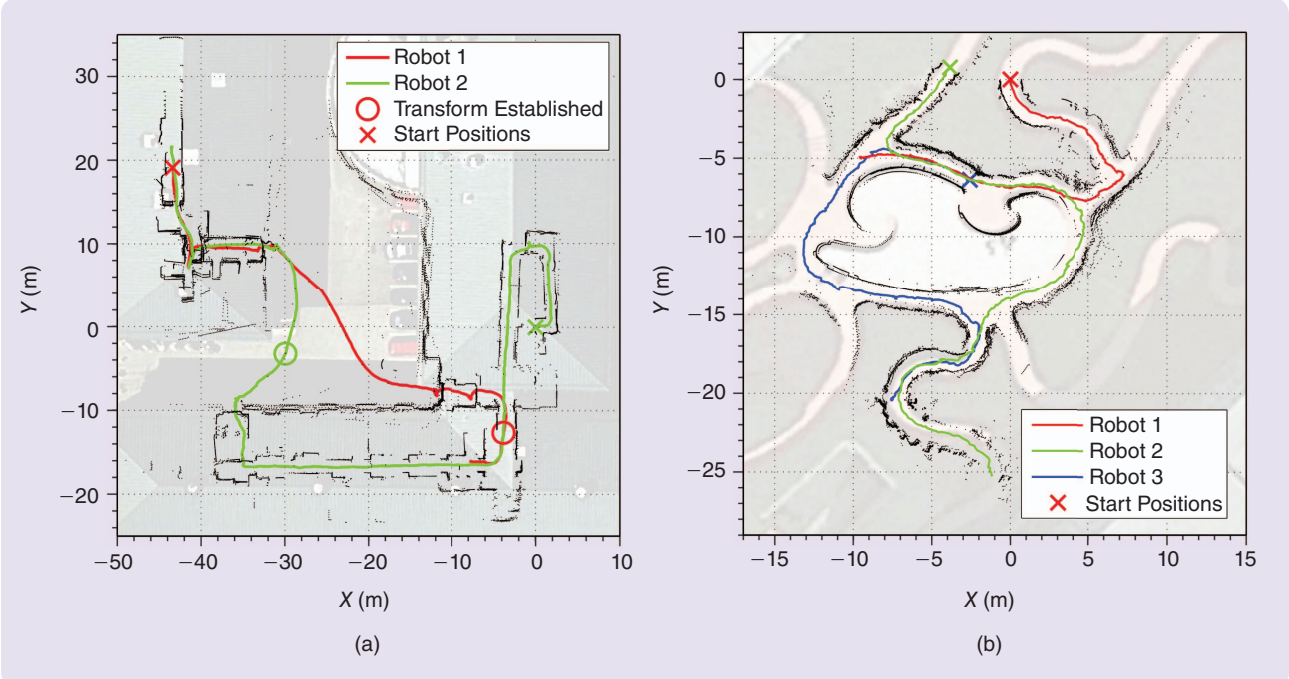


FIGURE 22 Data sets (a) **D5** and (b) **D6** with trajectories aligned in a common reference frame on top of satellite imagery.

were performed. **D4** involved three robots navigating from different starting points around a large indoor loop. In **D5**, two robots began in different buildings and then switched buildings (involving indoor-outdoor transitions) without directly observing one another. **D6** was performed with three robots that also never directly observed one another in an outdoor hedge maze. Table 5 contains information on the three large-scale data sets.

Figure 1(b) displays the trajectories of robots from data set **D4** after a common reference frame is computed. Individual robot trajectories in areas of the map that were not jointly explored by multiple robots contain drift, whereas portions of the trajectories near the central loop do not contain as much drift due to loop closures from multirobot correspondences.

Figure 22 shows the trajectories taken by robots from data sets **D5** and **D6**. Figure 22(a) is captured from the reference frame of the red robot in **D5**. Robot starting points are marked by crosses in the figure, and the points when robots settle on a transformation hypothesis are marked by circles. Since the two robots never directly observed one another, transformations were established after each robot found correspondences between its own sensor observations and past sensor observations from other robots. Figure 22(b) shows results from the outdoor data set **D6**, from the perspective of the blue robot. In contrast to other environments, this environment did not contain regular line and corner structures.

Timing Results

Figure 23 shows a breakdown of processing time by algorithmic component on one robot from a three-robot data set (**D2**). Pose-graph optimization [45], ICP matching, and EM optimization ("Pose-graph optimization," "Correspondence generation," and "EM" in Table 3, respectively) require a majority of the processing load. Hypothesis selection is efficient and only requires computation leading up to establishment of a common reference frame. Once the common reference frames are built, pose-graph optimization consumes the majority of the processing time. Since the cumulative time cost of the approach remains below 40 ms at all times, the implementation could be run at up to 25 Hz; 20 Hz is chosen so that the system can run in real time.

LIMITATIONS AND FUTURE WORK

While the described approach is capable of correctly inferring common reference frames and multirobot data association in various challenging scenarios, correctly identifying multirobot inlier correspondences becomes increasingly challenging as the robot poses develop significant drift over time. Incorporating robot pose uncertainty within the EM framework allows for identification of inlier multirobot correspondences, even in these challenging cases. However, the downside of this approach is that modifying the inlier covariance Σ_{inlier} could, theoretically, also result in identifying outlier correspondences as inliers, if the error in (11) in these correspondences is described well by the modified inlier covariance. This phenomenon has not been observed in any of the experiments performed. Nevertheless, this aspect deserves further investigation and is therefore a subject of future research.

Although the described approach is aimed at solving the multirobot SLAM problem in the presence of severe

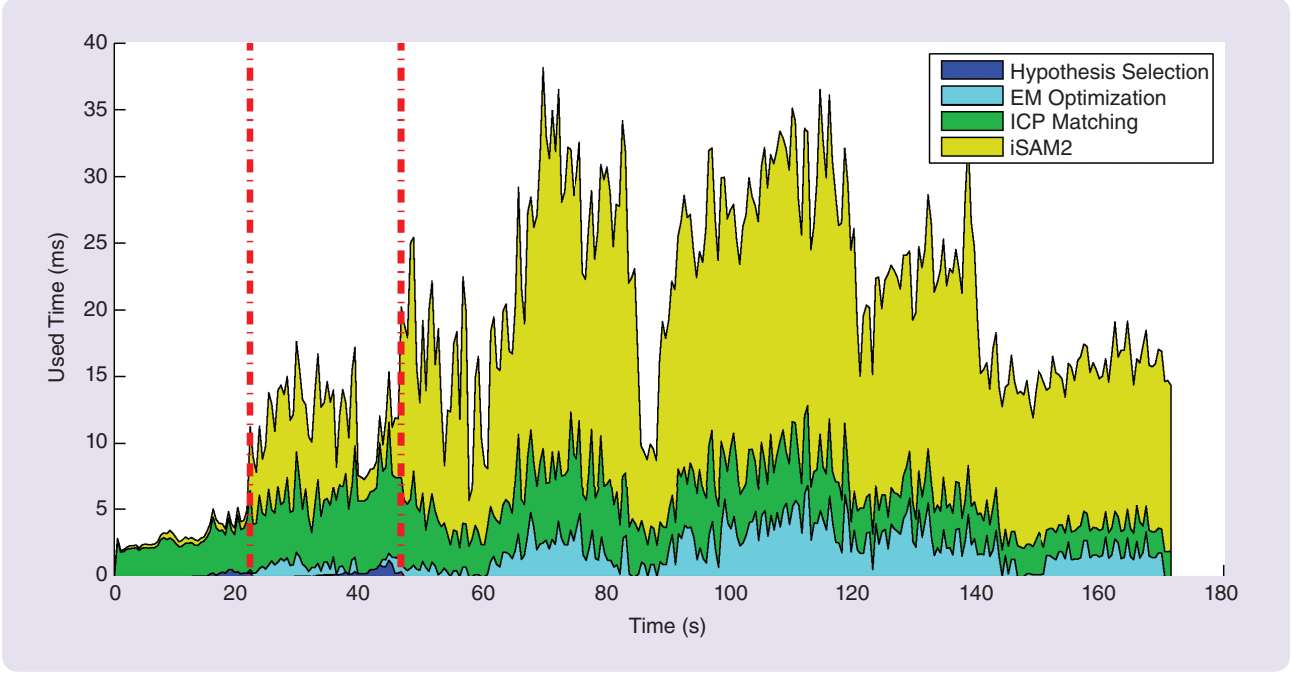


FIGURE 23 An algorithmic breakdown of processing times for data set **D2**. The two red lines mark the time stamps during which the robot establishes a common reference frame with another robot.

The underlying assumption in the presented approach is that robots operate in closed and finite environments and will eventually mutually observe a scene.

measurement aliasing, a higher-quality correspondence detection technique would enable more robust operation in challenging environments. The results were obtained using scan-by-scan ICP matching to generate correspondences. ICP frequently provides poor transformation estimates between laser scans due to local minima in the ICP optimization as well as sensor noise. Because of these failures, many correspondences generated by ICP are outliers. High outlier counts make it difficult to distinguish inliers and select hypotheses as the map size grows large. Although this article primarily focuses on the optimization ("back end") portion of the SLAM problem, any improvement to the correspondence generation strategy can improve the accuracy of the calculated relative transformations.

To handle the inaccuracies of ICP correspondences, a feature-based, correspondence-generation strategy that uses RANSAC [31] to match laser scan FLIRT feature descriptors [48] generated from the local and remote scans was described. A comparison between the two correspondence generation strategies is shown in Figure 24. More details of this feature-based approach is described in previous work [16]. The ICP strategy generates a high ratio of outliers while the feature-based approach generates few. The preliminary results of this feature-based method show that the approach could be in extended to larger and more complex environments.

The underlying assumption in the presented approach is that robots operate in closed and finite environments and will eventually mutually observe a scene (not necessarily at the same time). If the robots do not cross paths and in presence of perceptual aliasing, there will be (at least) one cluster that corresponds to consistent outliers that will compete with the null hypothesis, similar to the situation in Figure 15, although later on the robots do observe mutual scenes. If the robots never cross paths, the algorithm could break (an incorrect hypothesis will be chosen); however, typically, such cases violate the mentioned underlying assumption.

Finally, there remains significant flexibility in the choice of sensors and localization strategy run by each robot. Cameras and depth sensors are likely to provide richer correspondence estimates. In addition, a more robust localization strategy would diminish the drift caused by ICP odometry. Modifying sensors and localization strategies could extend the applications of the formulation.

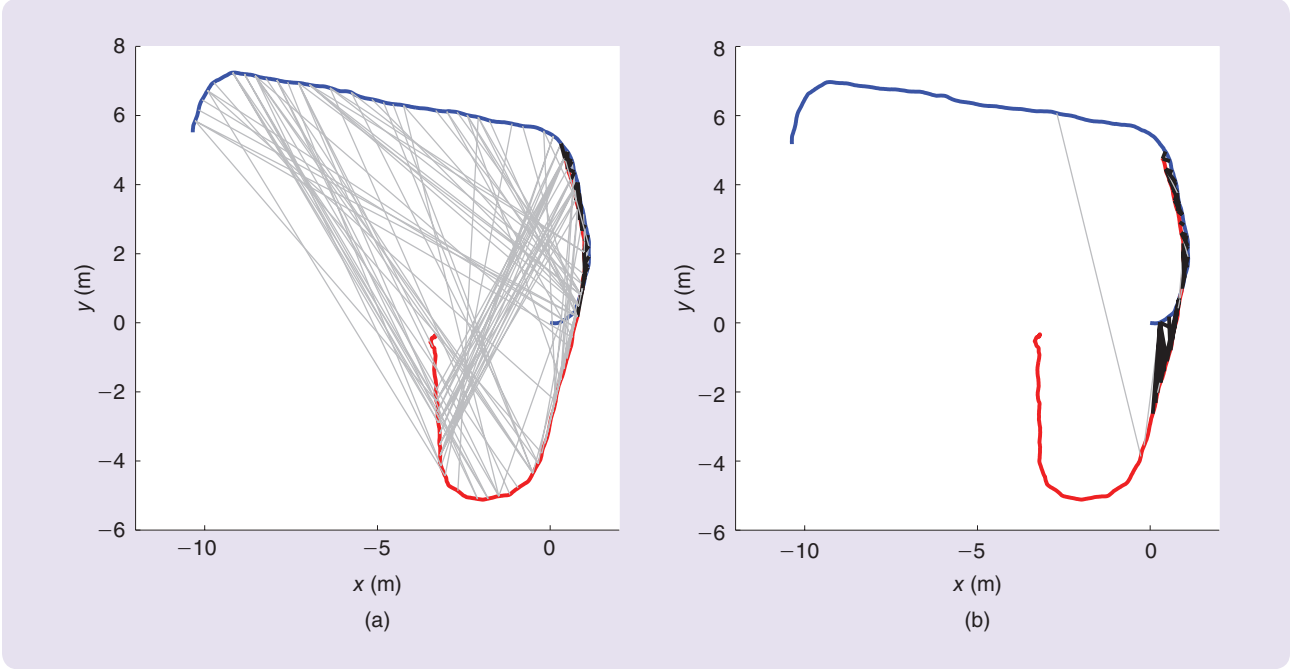


FIGURE 24 Correspondences generated using (a) an iterative closest point (ICP) and (b) the feature-based approach. The gray lines signify outliers, and black lines signify inliers. The feature-based strategy generates far fewer outlier correspondences.

The preliminary results of this feature-based method show that the approach could be extended to larger and more complex environments.

CONCLUSIONS

This article presented an approach for distributed and incremental localization and mapping by a group of collaborating robots that are initially unaware of each other's poses without assuming multirobot data association to be given. Under the assumption that the robots share informative observations with one another, an EM approach was developed to concurrently infer initial relative poses and multirobot data association. This approach is carried out by each robot in the group, starting from promising initial guesses and converging to several locally optimal solutions (referred to as hypotheses). Choosing the correct hypothesis is challenging in the incremental setting because of measurement aliasing and because there may be insufficient data to make this decision reliably. These challenges were addressed by developing a model-based selection approach that provides a probabilistically sound mechanism for choosing the most probable hypothesis, while modeling hypotheses prior using the CRP. Once a common reference frame between the robots is established, it becomes possible for each robot to perform a joint inference over its own and other robots' poses. Operation in the presence of significant drift in robot pose estimates, which is typical when operating in large-scale environments, was also considered. A modification of the approach was suggested to enable the identification of multirobot inlier correspondences even in these challenging cases. The method was evaluated in multirobot real-world indoor and outdoor experiments, which exhibited accurate estimation of common reference frames and correctly determined multirobot data association.

AUTHOR INFORMATION

Vadim Indelman (vadim.indelman@technion.ac.il) received the B.A. degree in computer science and the B.Sc. degree in aerospace engineering, both in 2002, and the Ph.D. degree in aerospace engineering in 2011, all from the Technion–Israel Institute of Technology. Between 2012 and 2014, he was a postdoctoral fellow at the Institute of Robotics and Intelligent Machines at the Georgia Institute of Technology. In July 2014, he assumed his present position of an assistant professor in aerospace engineering at the Technion. His research interests include autonomous navigation and mapping, distributed perception and information fusion, belief-space planning and active sensing, vision-aided navigation, and simultaneous localization and mapping. He can be contacted at the Department of Aerospace Engineering, Technion City, Haifa 32000, Israel.

Erik Nelson is a robotics M.S. student in the Robotics Institute at Carnegie Mellon University. He received the B.S. degree in materials engineering from California Polytechnic State University, San Luis Obispo, in 2013. His research focuses on robotic mapping, perception, and exploration. His current interests are in rapid autonomous exploration strategies for aerial robots.

Jing Dong received the B.Eng. degree in aerospace engineering from Tsinghua University, Beijing, China, in 2012. He is currently working toward the Ph.D. degree in computer science in the School of Interactive Computing, Georgia Institute of Technology. His research interests include visual simultaneous localization and mapping (SLAM), structure from motion, and multirobot SLAM.

Nathan Michael is an assistant research professor in the Robotics Institute at Carnegie Mellon University, Pittsburgh, Pennsylvania. He received the Ph.D. degree from the Department of Mechanical Engineering at the University of Pennsylvania in 2008 before transitioning into a research faculty position in the same department in 2010. His research focuses on enabling autonomous ground and aerial vehicles to robustly operate in uncertain environments with emphasis on robust state estimation, control, and cooperative autonomy.

Frank Dellaert is a professor in the School of Interactive Computing at the Georgia Institute of Technology. He obtained the Ph.D. degree in computer science from Carnegie Mellon University in 2001, the M.Sc. degree in electrical and computer engineering from Case Western Reserve University in 1995, and an electrical engineering degree from KU Leuven in 1989. His research focuses on probabilistic inference methods in robotics and computer vision, and he pioneered the Monte Carlo localization method for estimating and tracking the pose of robots. Most recently, he and his students have investigated three-dimensional reconstruction and mapping in large-scale environments by taking a graph-theoretic view.

REFERENCES

[1] N. Sunderhauf and P. Protzel, "Towards a robust back-end for pose graph slam," in *Proc. IEEE Int. Conf. Robotics Automation*, 2012, pp. 1254–1261.

[2] Y. Latif, C. D. C. Lerma, and J. Neira, "Robust loop closing over time," in *Proc. Robotics: Science Systems*, 2012, pp. 233–240.

[3] N. Sünderhauf and P. Protzel, "Switchable constraints vs. max-mixture models vs. rrr—A comparison of three approaches to robust pose graph slam," in *Proc. IEEE Int. Conf. Robotics Automation*, 2013, pp. 5198–5203.

[4] G. H. Lee, F. Fraundorfer, and M. Pollefeys, "Robust pose-graph loop-closures with expectation-maximization," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 2013, pp. 556–563.
[5] E. Olson and P. Agarwal, "Inference on networks of mixtures for robust robot mapping," *Int. J. Robot. Res.*, vol. 32, no. 7, pp. 826–840, 2013.
[6] A. Cunningham, K. Wurm, W. Burgard, and F. Dellaert, "Fully distributed scalable smoothing and mapping with robust multi-robot data association," in *Proc. IEEE Int. Conf. Robotics Automation*, Saint Paul, MN, 2012, pp. 1050–4729.
[7] E. Montijano, R. Aragues, and C. Sagues, "Distributed data association in robotic networks with cameras and limited communications," *IEEE Trans. Robot.*, vol. 29, no. 6, pp. 1408–1423, 2013.
[8] J. Knuth and P. Barooah, "Outlier rejection for pose graph optimization," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 2013, pp. 5198–5203.
[9] 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 *Proc. IEEE Int. Conf. Robotics Automation*, 2014, pp. 593–600.
[10] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Mag.*, vol. 13, no. 6, pp. 47–60, 1996.
[11] K. P. Burnham and D. R. Anderson, *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. Berlin, Germany: Springer-Verlag, 2002.
[12] D. Blei, T. Griffiths, M. Jordan, and J. Tenenbaum, "Hierarchical topic models and the nested Chinese restaurant process," *Adv. Neural Inform. Process. Syst.*, vol. 16, no. 16, pp. 17–24, 2003.
[13] Y. W. Teh, M. Jordan, M. Beal, and D. Blei, "Sharing clusters among related groups: Hierarchical Dirichlet processes," in *Proc. Advances Neural Information Processing Systems*, 2004.
[14] V. Indelman, N. Michael, and F. Dellaert, "Incremental distributed robust inference from arbitrary robot poses via em and model selection," in *Proc. RSS Workshop Distributed Control Estimation Robotic Vehicle Networks*, July 2014, pp. 1–4.
[15] E. Nelson, V. Indelman, N. Michael, and F. Dellaert, "An experimental study of robust distributed multi-robot data association from arbitrary poses," in *Experimental Robotics, The 14th International Symposium on Experimental Robotics*. New York: Springer: 2016, pp. 323–338.
[16] J. Dong, E. Nelson, V. Indelman, N. Michael, and F. Dellaert, "Distributed real-time cooperative localization and mapping using an uncertainty-aware expectation maximization approach," in *Proc. IEEE Int. Conf. Robotics Automation*, 2015, pp. 5807–5814.
[17] J. Fenwick, P. Newman, and J. Leonard, "Cooperative concurrent mapping and localization," in *Proc. IEEE Int. Conf. Robotics Automation*, 2002, vol. 2, pp. 1810–1817.
[18] S. Roumeliotis and G. Bekey, "Distributed multirobot localization," *IEEE Trans. Robot. Automat.*, vol. 18, no. 5, pp. 781–785, Oct. 2002.
[19] A. Howard, "Multi-robot simultaneous localization and mapping using particle filters," *Int. J. Roboti. Res.*, vol. 25, no. 12, pp. 1243–1256, 2006.
[20] A. Howard, G. S. Sukhatme, and M. J. Matarić, "Multi-robot mapping using manifold representations," *Proc. IEEE Special Issue Multi-Robot Syst.*, vol. 94, pp. 1360–1369, July 2006.
[21] B. Kim, M. Kaess, L. Fletcher, J. Leonard, A. Bachrach, N. Roy, and S. Teller, "Multiple relative pose graphs for robust cooperative mapping," in *Proc. IEEE Int. Conf. Robotics Automation*, Anchorage, AK, May 2010, pp. 3185–3192.
[22] T. Bailey, M. Bryson, H. Mu, J. Vial, L. McCalman, and H. Durrant-Whyte, "Decentralised cooperative localisation for heterogeneous teams of mobile robots," in *Proc. IEEE Int. Conf. Robotics Automation*, 2011, pp. 2859–2865.
[23] V. Indelman, P. Gurfil, E. Rivlin, and H. Rotstein, "Distributed vision-aided cooperative localization and navigation based on three-view geometry," *Robot. Auton. Syst.*, vol. 60, pp. 822–840, June 2012.
[24] V. Indelman, P. Gurfil, E. Rivlin, and H. Rotstein, "Graph-based distributed cooperative navigation for a general multi-robot measurement model," *Int. J. Robot. Res.*, vol. 31, Aug. 2012.
[25] J. M. Walls and R. M. Eustice, "An exact decentralized cooperative navigation algorithm for acoustically networked underwater vehicles with robustness to faulty communication: Theory and experiment," in *Proc. Robotics: Science Systems Conf.*, 2013.
[26] X. Zhou and S. Roumeliotis, "Multi-robot SLAM with unknown initial correspondence: The robot rendezvous case," in *IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 2006, pp. 1785–1792.
[27] L. Andersson and J. Nygards, "C-SAM: Multi-robot SLAM using square root information smoothing," in *Proc. IEEE Int. Conf. Robotics Automation*, 2008, pp. 2798–2805.
[28] 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 *Proc. IEEE Int. Conf. Robotics Automation*, 2010, pp. 243–249.
[29] A. Franchi, G. Oriolo, and P. Stegagno, "Mutual localization in multi-robot systems using anonymous relative measurements," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 81–95, 2013.
[30] 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," *Robot. Automat. Mag.*, vol. 13, no. 3, pp. 53–62, 2006.
[31] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[32] A. Howard, L. E. Parker, and G. S. Sukhatme, "Experiments with a large heterogeneous mobile robot team: Exploration, mapping, deployment and detection," *Int. J. Robot. Res.*, vol. 25, no. 5–6, pp. 431–447, 2006.
[33] S. Carpin, "Fast and accurate map merging for multi-robot systems," *Auton. Robots*, vol. 25, no. 3, pp. 305–316, 2008.
[34] S. Saeedi, L. Paull, M. Trentini, M. Seto, and H. Li, "Group mapping: A topological approach to map merging for multiple robots," *Robot. Automat. Mag.*, vol. 21, no. 2, pp. 60–72, 2014.
[35] M. Cummins and P. Newman, "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," *Int. J. Robot. Res.*, vol. 27, pp. 647–665, June 2008.
[36] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2nd ed., 2004.
[37] A. Cunningham, V. Indelman, and F. Dellaert, "DDF-SAM 2.0: Consistent distributed smoothing and mapping," in *Proc. IEEE Int. Conf. Robotics Automation*, Karlsruhe, Germany, May 2013, pp. 5220–5227.
[38] H. Durrant-Whyte and M. Stevens, "Data fusion in decentralized sensing networks," in *Proc. 4th Int. Conf. Information Fusion*, 2001, pp. 302–307.
[39] A. Bahr, M. Walter, and J. Leonard, "Consistent cooperative localization," in *Proc. IEEE Int. Conf. Robotics Automation*, May 2009, pp. 3415–3422.
[40] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Auton. Robots*, vol. 4, no. 4, pp. 333–349, Apr. 1997.
[41] D. Chetverikov, D. Stepanov, and P. Krsek, "The trimmed iterative closest point algorithm," in *Proc. Int. Conf. Pattern Recognition*, 2002, pp. 545–548.
[42] T. Minka. (1998, Nov.). Expectation-maximization as lower bound maximization. [Online]. Available: http://research.microsoft.com/en-us/um/people/minka/papers/em.html
[43] A. Ranganathan and F. Dellaert, "Online probabilistic topological mapping," *Int. J. Robot. Res.*, vol. 30, no. 6, pp. 755–771, 2011.
[44] J. Nieto, T. Bailey, and E. Nebot, "Recursive scan-matching slam," *Robot. Auton. Syst.*, vol. 55, no. 1, pp. 39–49, 2007.
[45] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *Int. J. Robot. Res.*, vol. 31, pp. 217–236, Feb. 2012.
[46] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," Tech. Rep. GT-RIMCP&R-2012-002, Inst. Robotics Intelligent Machines, Georgia Inst. of Technology, 2012.
[47] S. Shen, N. Michael, and V. Kumar, "Autonomous multi-floor indoor navigation with a computationally constrained MAV," in *Proc. IEEE Int. Conf. Robotics Automation*, 2011, pp. 20–25.
[48] G. D. Tipaldi and K. O. Arras, "Flirt-interest regions for 2D range data," in *Proc. IEEE Int. Conf. Robotics Automation*, 2010, pp. 3616–3622.
[49] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
[50] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G^2o: A general framework for graph optimization," in *Proc. IEEE Int. Conf. Robotics Automation*, Shanghai, China, May 2011, pp. 3607–3613.
[51] M. Kaess and F. Dellaert, "Covariance recovery from a square root information matrix for data association," *Robot. Auton. Syst.*, vol. 57, no. 12, pp. 1198–1210, 2009.