Pairwise Consistent Measurement Set Maximization for Robust Multi-robot Map Merging

Joshua G. Mangelson, Derrick Dominic, Ryan M. Eustice, and Ram Vasudevan

***Abstract*— This paper reports on a method for robust selection of inter-map loop closures in multi-robot simultaneous localization and mapping (SLAM). Existing robust SLAM methods assume a good initialization or an "odometry backbone" to classify inlier and outlier loop closures. In the multi-robot case, these assumptions do not always hold. This paper presents an algorithm called Pairwise Consistency Maximization (PCM) that estimates the largest *pairwise internally consistent set* of measurements. Finding the largest pairwise internally consistent set can be transformed into an instance of the maximum clique problem from graph theory, and by leveraging the associated literature it can be solved in real-time. This paper evaluates how well PCM approximates the combinatorial gold standard using simulated data. It also evaluates the performance of PCM on synthetic and real-world data sets in comparison with DCS, SCGP, and RANSAC, and shows that PCM significantly outperforms these methods.**

I. Introduction

In multi-agent simultaneous localization and mapping (SLAM), multiple robots collect measurements about their own trajectories, the environment, and possibly other robots. To generate an accurate map of the environment, it is necessary to estimate both the local trajectories of the robots as well as the relative offsets (translation and orientation) between the trajectories.

In pose graph SLAM, the map estimation problem is formulated as a factor graph consisting of pose node variables and factor nodes. We often formulate the single robot pose graph SLAM problem as the maximum likelihood estimate (MLE) of the time discretized robot trajectory given odometric and loop closure measurements [1]. Assuming independent measurements and additive Gaussian noise in the measurement and robot process models, this becomes a non-linear, weighted least squares problem and can be solved quickly using available solvers [2, 3].

However, accurately determining large loop closure factors is a difficult problem, and least squares can be susceptible to outliers. A variety of methods have been proposed to enhance robustness by disabling factors with high residual error [4–7], but they either depend on having an odometry backbone of trusted measurements to judge against or a good prior on the relative pose of the vehicles [8]. In multi-agent SLAM, this odometry backbone is non-existent, and a prior on relative pose may not exist.

The use of multiple vehicles enables increased scalability and efficiency in the mapping process. However, to accurately fuse their maps, the vehicles must estimate their position with respect to one another. This process, very similar to the loop closure problem in single robot SLAM, requires accurately selecting a set of consistent, perception-derived pose measurements relating the relative position and orientation of the two vehicles and their maps. In this paper, we present a tractable and robust way of selecting which measurements to trust.

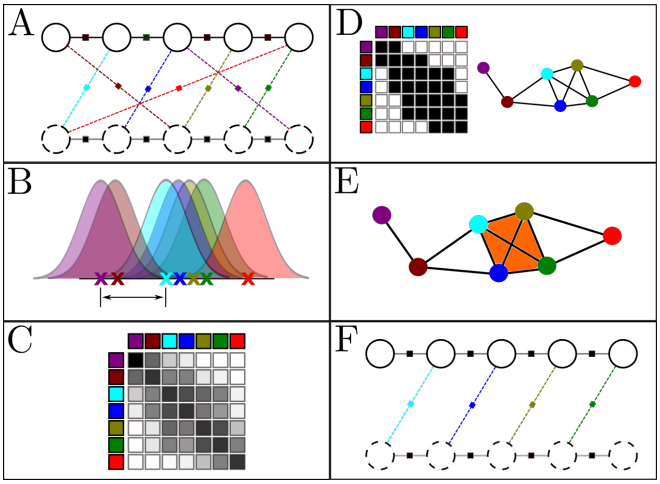

Fig. 1. An illustration of the Pairwise Consistency Maximization (PCM) algorithm for selecting consistent inter-map loop closures measurements. (A) Given two independently derived pose graphs (shown in white and black in step A) and a set of potential loop closures between them (shown by colored, dotted lines), our goal is to determine which of these inter-robot loop closures should be trusted. (B) Using a consistency metric such as Mahalanobis distance, we calculate the consistency of each pairwise combination of measurements. (C) We store these pairwise consistency values in a matrix where each element corresponds to the consistency of a pair of measurements. (D) We can transform this matrix into the adjacency matrix for a *consistency graph* by thresholding the consistency and making it symmetric using the maximum consistency when associated elements across the diagonal have differing consistency values. Each node in this graph represents a measurement and edges denote consistency between measurements. Cliques in this graph are *pairwise internally consistent sets*. (E) Finding the maximum clique represents finding the largest pairwise internally consistent set. (F) After determining the largest consistent set, we can robustly merge the two pose graphs using only the consistent inter-map loop closures, allowing us to reject false measurements.

Rather than attempt to classify measurements as inliers and outliers, we find the largest consistent set of inter-robot relative pose measurements. We first formulate this problem as a combinatorial optimization problem, which turns out to be an instance of the NP-complete maximum clique problem. We then present a method from graph theory literature that finds the optimal solution for moderately sized problems (including many multi-robot map merging scenarios), as well as a heuristic that can be used to approximate it for larger numbers of measurements.

*This work was supported by the Office of Naval Research under award N00014-16-1-2102.

J. Mangelson, D. Dominic, R. Eustice, and R. Vasudevan are at the University of Michigan, Ann Arbor, MI 48109, USA.

{mangelso, ddominic, eustice, ramv}@umich.edu.

Our contibutions include the following: *i*) In §IV, a novel definition of *pairwise internally consistent sets* and the formulation of the robust multi-vehicle map merging problem as a combinatorial problem that seeks to find the maximum cardinality set of internally consistent measurements. *ii*) In §V, a method for transforming the PCM problem into the equivalent problem of finding the maximal clique of a *consistency graph*, for which several algorithms exist.

The remainder of this paper is organized as follows. In §II, we cover related work. In §III, we present the general formulation of the multi-robot pose graph SLAM problem. In §VI, we outline the steps needed to integrate PCM into a map merging system. In §VII, we evaluate PCM on synthetic and real-world datasets and show that PCM outperforms the state-of-the-art methods in selecting consistent measurements and estimating the merged maps. Finally in §VIII, we conclude.

II. Related Work

Current methods for identifying loop closures often severely limit the number of accepted measurements by setting high likelihood thresholds in an attempt to filter out false positives [9]. In pose graph SLAM, inconsistent measurements must be filtered or they may distort the estimated graph.

The detection of outliers is an important problem in both mobile robotics and computer vision and a variety of methods have been proposed to address it. In computer vision, the RANSAC algorithm seeks to determine inlier/outlier sets by iteratively fitting models to samples of the data and counting the number of inliers [10]. However, RANSAC relies on the fact that a point can be designated as an inlier or outlier by measuring its likelihood given a proposed model. However, in multi-robot SLAM, there is no unique model we can use to determine if a constraint is an inlier. Rather, we need to know if a measurement is consistent with all the other measurements being considered as inliers. In addition, RANSAC and many of its derivatives are sensitive to parameter values that can be difficult to tune.

JCBB is a method that selects measurements by seeking to determine the maximum jointly compatible set [11]. In the multi-robot map merging problem, however, performing JCBB can be difficult because doing so would require solving the graph for a combinatorial number of measurement combinations to evaluate the likelihood of each measurement given each combination of the other measurements.

There have been several outlier detection methods derived specifically for pose graph SLAM. Switchable constraints [4] proposes the use of of switchable error factors in the SLAM back-end solver that enable pose constraints with high residual error to be "turned off" so that they no longer affect the solution. Dynamic covariance scaling [6] (DCS) builds on [4] and uses scaling of measurement covariance to more smoothly turn on and off measurement links between poses. In [5] Max-Mixtures uses mixtures of Gaussians to model multiple possible data modes. Realizing, Reversing, Recovering [7] iteratively tries to find a set of relative pose measurements that are consistent within themselves. However, each of these methods is designed with a single robot pose graph in mind and fails without a good initialization of relative pose [4–6] or a trusted odometry backbone [7].

Recently, several papers have been published directly attempting to solve the "perceptual aliasing" problem of map merging. Perceptual aliasing occurs when multiple locations in the environment look similar enough that perceptual sensors and methods mistake them for the same location. The method presented in [12, 13] uses expectation maximization to estimate the inliers and outliers in a set of measurements that constrain the relative poses of multiple robots performing cooperative mapping. Our proposed method uses many similar concepts to those presented in [12], but our solution finds/estimates the maximum cardinality consistent set of inter-robot measurements instead of iteratively refining an initial guess. Generalized graph SLAM [8] handles multi-hypothesis or outlier factors that could be derived from perceptual aliasing by attempting to find the minimum ambiguity spanning tree for the graph. However, their method does not attempt to enforce consistency of measurements and, in the multi-robot case, may select an incorrect measurement if it happens to have the lowest uncertainty.

The two most similar methods to ours are [14] and [15]. Our method follows the same approach as [14] in that we both attempt to find the maximal set of measurements that are consistent with one another. However, their work has only been extended to non-linear measurements in the case of planar SLAM. Single Cluster Spectral Graph Partitioning (SCGP) presented in [15, 16]. SCGP [16] attempts to determine a single cluster of measurements, given a similarity matrix, by thresholding the elements of the largest eigenvector. [15] uses a similar consistency metric to ours, but they do not enforce pairwise internal set consistency as we do.

III. Problem Formulation

In our factor graph formulation of SLAM, we denote time discretized versions of the the robot trajectory by $\mathbf{x}_i \in \mathrm{SE}(2)$ or $\mathrm{SE}(3)$. The factors in the graph are derived from the measurements observed by the robot and penalize estimates of the trajectory that make the observed measurement unlikely. We denote measurements that relate the variables $\mathbf{x}_i$ and $\mathbf{x}_j$ by $\mathbf{z}_{ij}$ and call them odometric measurements if i and j are consecutive and loop closure measurements if i and j are non-consecutive in time. The goal of pose graph SLAM is, then, to estimate the most likely value of each pose variable $\mathbf{x}_i$ given the factor measurements $\mathbf{z}_{ij}$. We can formulate the single robot pose graph SLAM problem as the MLE

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmax}}\, P(\mathbf{Z}|\mathbf{X}). \tag{1}$$

where, $\mathbf{X}$ is the set of all pose variables $\mathbf{x}_i$, and $\mathbf{Z}$ is the set of all relative pose measurements robot measurements $\mathbf{z}_{ij}$.

In multi-robot SLAM, we also need to estimate the relative transformation between the local coordinate frames of the respective robots. We adopt the method presented by Kim et al. [17], which proposes the use of an anchor node for each trajectory that encodes the pose of the vehicle's

local coordinate frame with respect to some global reference frame. We denote the homogeneous transformation matrix representing this offset by T_a^g and represent measurements relating cross-trajectory poses by $\mathbf{z}_{ij}^{ab}$, where a and b are robot IDs and i and j respectively denote which poses on robots a and b are being related. T_a^g is an element of $\mathrm{SE}(2)$ or $\mathrm{SE}(3)$. $\mathbf{z}_{ij}^{ab}$ is also often an element of $\mathrm{SE}(2)$ or $\mathrm{SE}(3)$ but can be a function of this transformation in general.

In the case of two robots, the SLAM estimation problem becomes

$$\hat{\mathbf{X}}, \hat{\mathbf{T}}^g = \underset{\mathbf{X}, \mathbf{T}^g}{\operatorname{argmax}} P(\mathbf{Z}^a, \mathbf{Z}^b, \mathbf{Z}^{ab} | \mathbf{X}, \mathbf{T}^g), \tag{2}$$

where, $\mathbf{X}$ now represents the trajectories of both robots, $\mathbf{Z}^{ab}$ represents the set of all cross-trajectory measurements, $\mathbf{Z}^r$ represents the set of measurements local to robot r, and $\mathbf{T}^g = \{\mathbf{T}_a^g, \mathbf{T}_b^g\}$. This problem can be treated as weighted, non-linear least squares and can be solved efficiently using an array of specialized optimization libraries.

Existing methods do a good job of handling outlier measurements in the local measurement sets $\mathbf{Z}^a$ and $\mathbf{Z}^b$, but not in the inter-robot set $\mathbf{Z}^{ab}$ since no prior estimate of the initial transformation between the robot coordinate frames exists in general. The focus of this paper is on selecting a subset of the measurements in the inter-robot set $\mathbf{Z}^{ab}$ that we can trust. The next section outlines our approach for doing so.

IV. Pairwise Consistency Maximization

In this section, we first define a novel notion of consistency and then we use that notion to formulate the selection of inter-robot loop closure measurements as a combinatorial optimization problem that finds the largest consistent set.

A. Pairwise Consistency

Directly determining if a measurement is an inlier or outlier from the graph itself is unobservable [14]. Thus, instead of trying to classify inlier versus. outlier, we attempt to determine the maximum subset of measurements that are internally pairwise consistent:

Definition 1: *A set of measurements $\widetilde{\mathbf{Z}}$ is **pairwise internally consistent** with respect to a consistency metric C and the threshold γ if*

$$C(\mathbf{z}_i, \mathbf{z}_j) \leq \gamma, \quad \forall \quad \mathbf{z}_i, \mathbf{z}_j \in \widetilde{\mathbf{Z}} \tag{3}$$

where, C is a function measuring the consistency of measurements $\mathbf{z}_i$ and $\mathbf{z}_j$, and γ is chosen a priori.

This definition of consistency requires that every measurement in the set be consistent with every other measurement in the set with respect to C and γ.

There are a variety of potential choices of a consistency metric depending on the measurement model and the state being observed. For the remainder of this paper, however, we assume that all inter-robot measurements are relative pose measurements with full degrees of freedom and use the following metric based on [15]:

$$C(\mathbf{z}_{ik}^{ab}, \mathbf{z}_{jl}^{ab}) = \left|\left|(\ominus \mathbf{z}_{ik}^{ab}) \oplus \hat{\mathbf{x}}_{ij}^a \oplus \mathbf{z}_{jl}^{ab} \oplus \hat{\mathbf{x}}_{lk}^b\right|\right|_\Sigma \triangleq ||\epsilon_{ikjl}||_{\Sigma_{ikjl}} \tag{4}$$

where, we have adopted the notation of [18] to denote pose composition using $\oplus$ and inversion using $\ominus$, $||\cdot||_\Sigma$ signifies the Mahalanobis distance, and the variables $\hat{\mathbf{x}}_{ij}^a$ and $\hat{\mathbf{x}}_{lk}^b$ are the current relative pose estimates of the associated poses corresponding to inter-robot measurements $\mathbf{z}_{ik}^{ab}$ and $\mathbf{z}_{jl}^{ab}$.

This choice of metric is useful because it is both easy to compute and follows a chi-squared distribution, giving us a strategy to select the threshold γ without knowledge of the specific dataset. The composition inside the norm of (4) evaluates the pose transformation around a loop and should evaluate to the identity transformation in the case of no noise [15]. With Gaussian noise, this normalized squared error follows a chi-squared distribution with degree of freedom equal to the number of degrees of freedom in our state variable. By setting γ accordingly, we can determine if the measurements $\mathbf{z}_{ik}^{ab}$ and $\mathbf{z}_{jl}^{ab}$ are consistent with one another.

It should also be noted that pairwise consistency does not necessarily signify full joint consistency. It is possible that a set of measurements can be pairwise internally consistent but not jointly consistent. However, checking full joint consistency is an exponential operation and requires possibly checking every combination of measurements to evaluate their consistency. Finding the maximum cardinality pairwise consistent set is also exponential, but by formulating the problem in this way, we can leverage a body of literature on the maximum clique problem in graph theory that can find or estimate the solution efficiently. In addition, in practice we observed that testing for pairwise consistency was restrictive enough to filter inconsistent measurements from typical pose graphs with full degree of freedom measurements.

B. The Maximal Cardinality Pairwise Consistent Set

Having this definition of pairwise internal consistency allows us to restrict our algorithm to only consider sets of measurements that are pairwise internally consistent; however, due to perceptual aliasing, we may end up with multiple subsets that are pairwise internally consistent. We need to find a way to select between these possible subsets.

The underlying assumption of our method is based on the following two initial assumptions:

Assumption 1: *The pose graphs are derived from multiple robots or the same robot in multiple sessions exploring the same environment.*

Assumption 2: *The inter-robot relative pose measurements are derived from observations of that environment and the system used to derive them is not biased toward selecting incorrect measurements over correct ones.*

These assumptions fit a large number of multi-robot mapping situations and are reasonable even in perceptually aliased environments whenever a place recognition system does not systematically select the perceptually aliased measurement over the correct ones.

If the above conditions are met than the following can also be safely assumed:

Assumption 3: *As the number of relative pose measurements increases, the number of measurements in the correct consistent subset will grow larger than those in the perceptually aliased consistent subsets.*

Our goal is, then, to efficiently find the largest consistent subset of $\mathbf{Z}^{ab}$, which we denote by $\mathbf{Z}^*$.

To formalize this, we introduce a binary switch variable, s_u, for each constraint in the set $\mathbf{Z}^{ab}$ and let s_u take on the value 1 if the measurement is contained in the chosen subset and 0 otherwise. Note that there is a single s_u for each measurement $\mathbf{z}_{ij}^{ab} \in \mathbf{Z}^{ab}$; however, for simplicity of notation, we now re-number them with the single index u and denote the corresponding measurement $\mathbf{z}_{ij}^{ab}$ by $\mathbf{z}_u$. Letting $\mathbf{S}$ be the vector containing all s_u, our goal is to find the solution, $\mathbf{S}^*$, to the following optimization problem:

$$\begin{aligned} \mathbf{S}^* &= \underset{\mathbf{S}\in\{0,1\}^m}{\operatorname{argmax}} \ \|\mathbf{S}\|_0 \\ \text{s.t.} \ &||\epsilon_{uv}||_{\Sigma_{uv}} \, s_u s_v \leq \gamma, \end{aligned} \tag{5}$$

where, m is the number of measurements in $\mathbf{Z}^{ab}$, $\mathbf{z}_u$ is the measurement corresponding to s_u, ϵ_{uv} is the associated error term corresponding to measurements $\mathbf{z}_u$ and $\mathbf{z}_v$, and Σ_{uv} is the covariance matrix associated with the error ϵ_{uv}. We refer to this as the PCM problem.

Once found, we can use $\mathbf{S}^*$ to index into $\mathbf{Z}^{ab}$ and get $\mathbf{Z}^*$. This consistent subset of the measurements can then be plugged into any of the existing non-linear least squares based solvers to merge the individual robot maps into a common reference frame. In the next section, we show how this problem can be reformulated into an equivilent problem that has been well studied.

V. Solving PCM via Maximum Clique

In this section, we describe how to solve the PCM problem. The goal of PCM to determine the largest subset of the measurements $\mathbf{Z}^{ab}$ that are pairwise internally consistent. This pairwise consistency is enforced by the n^2 constraints listed in (5). It is important to note that the norm on the left-hand side of the constraints does not contain any of the decision variables s_i. These distance measures can be calculated in pre-processing, as will be explained in §VI, and combined into a matrix of consistency measures $\mathbf{Q}$, where each element $[\mathbf{Q}]_{uv} = q_{uv} = ||\epsilon_{uv}||_{\Sigma_{uv}}$, corresponds to the consistency of measurement $\mathbf{z}_u$ and $\mathbf{z}_v$. This process is depicted in steps B and C in Fig. I.

We'll now introduce the concept of a *consistency graph*.

Definition 2: *A **consistency graph** is a graph $G = \{V, \mathcal{E}\}$ where each vertex $v \in V$ represents a measurement and each edge $e \in \mathcal{E}$ denotes consistency of the vertices it connects.*

We can transform the matrix of consistency measures $\mathbf{Q}$ into the adjacency matrix for a consistency graph if we threshold it by γ and make it symmetric by requiring that both q_{uv} and q_{vu} be less than or equal to γ to insert an edge into the graph. An example adjacency matrix and consistency graph are shown in step D of Fig. I.

A *clique* in graph theory is defined as a subset of vertices in which every pair of vertices has an edge between them and the *maximum clique* is the largest such subset of nodes in the graph. A clique of the consistency graph corresponds to a *pairwise internally consistent set* of measurements because every measurement is pairwise consistent with every other measurement in the set. Thus, the solution to the problem defined in (5) is the maximum clique of the consistency graph (see step E of Fig. I).

In graph theory, the problem of finding the maximum clique for a given graph is called the maximum clique problem and is an NP-hard problem [19]. The maximum clique problem is also hard to approximate [20, 21], meaning that finding a solution arbitrarily close to the true solution is also NP-hard. Dozens of potential solutions have been proposed, each of which can be classified as either an exact or a heuristic algorithm. All of the exact algorithms are exponential in complexity and are usually based on branch and bound, while the heuristic algorithms often try to exploit some type of structure in the problem, making them faster, but not guaranteeing the optimal solution [19].

In 2015, Pattabiraman et al. [22] proposed a method that aggressively prunes the search tree and is able to find maximum clique solutions for large sparse graphs relatively quickly. They present both an exact algorithm as well as a heuristic version that can be used when the exact algorithm becomes intractable. Though our method could theoretically use any one of the proposed maximum clique algorithms, we selected the one proposed in [22] because of its simplicity, parallelizablity, and open source implementation.

VI. Robust Multi-Robot Map Merging

Integrating PCM into a map merging system consists of four steps: *i*) Individual map generation *ii*) Consistency calculation *iii*) Pairwise consistency maximization *iv*) Constraint insertion and full map generation

The individual map generation step consists of solving for an estimate of the individual pose graphs. This can be carried out by any pose graph SLAM-based method as long as it is possible to estimate the marginal and cross-covariances of nodes and sets of nodes after a solution has been found. This is important for estimating the Mahalanobis distance matrix.

Using the consistency metric we have selected, the first step in estimating the Mahalanobis distance matrix is to extract the necessary marginal covariances. As explained in §IV-A, we take each pair of potential, inter-robot measurements and trace a loop using the transformations from the underlying graph. To trace this loop, we extract the covariance for every pair of nodes on a given local map that are associated with an inter-robot constraint. Then, using the methods described in [18], we calculate the "tail-to-tail" between these nodes allowing us to estimate $\hat{\mathbf{x}}_{ij}^a$ and $\hat{\mathbf{x}}_{lk}^b$ and their associated covariances. Finally, we use (4) to estimate the distances $||\epsilon_{ijkl}||_{\Sigma_{ijkl}}$ needed in (5). $\Sigma_{ijkl} = J\Sigma J^\top$, where J is equal to the Jacobian of ϵ_{ijkl} with respect to the two measurements and the two estimated local trajectory components evaluated at their means. Σ is equal to the associated block diagonal covariance matrix built up of the covariances of the measurements and pose estimates.

In practice, this calculation is the bottleneck in processing for our algorithm; however, by using analytical Jacobians, parallelization, and incremental updates, we can significantly reduce this calculation time. For more information on calculating the Mahalanobis distance see [18].

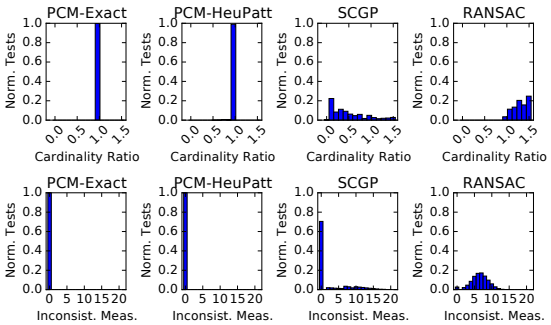

Fig. 2. Histograms that evaluate how well PCM, SCGP [16], and RANSAC [10] approximate the combinatorial maximum pairwise consistent set in (5). The first row of histogram plots shows the size of the measurement set as compared to the maximum consistent set size. The second row of histograms shows the number of inconsistent pairs returned with respect to the set γ threshold on Mahalanobis distance.

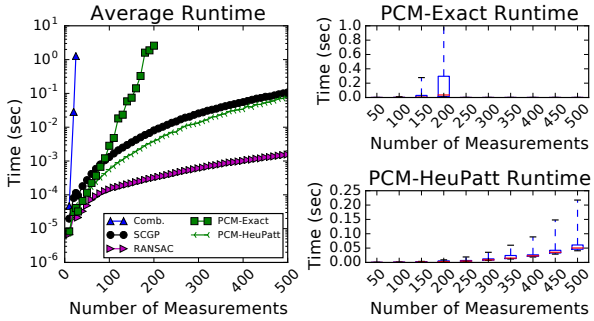

Fig. 3. A plot of the evaluation times of the different methods versus the number of measurements being tested. The combinatorial solution takes exponential time and PCM-Exact takes exponential time in the worst case, while the other methods are polynomial in the number of measurements. (This excludes the time to estimate the distance matrix $\mathbf{Q}$, which is required for all methods.)

Once the matrix has been calculated, we convert it to an adjacency matrix and solve the maximum clique problem as explained in §V to determine the set of constraints that should be enabled. Finally, we re-solve the graph using only those inter-robot constraints. In practice, it helped when optimizing over $\mathrm{SE}(3)$ to incrementally re-build the graph running a batch optimization every few hundred nodes, while for $\mathrm{SE}(2)$ adding the inter-robot factors directly to the already existing graph and running a batch optimization was sufficient.

VII. Evaluation

In this section, we evaluate the performance of PCM on a variety of synthetic and real-world data-sets. For comparison, we implemented single cluster graph partitioning (SCGP) [16], dynamic covariance scaling (DCS) [6], and random sample consensus (RANSAC) [10].

We implemented SCGP as described in [16], with the exception of using an off the shelf eigen-factorization library as opposed to the power method for simplicity. We implemented DCS as described in the original paper [6] with $\phi = 5$.

We implemented RANSAC by iteratively selecting a single, random inter-map measurement and evaluating the likelihood of the other measurements given the model estimated from the sampled measurement. Because the processing time for this evaluation is so low (given that the Mahalanobis distance evaluations were performed in pre-processing), we exhaustively iterate through all the measurements and evaluate the likelihood of the other measurements with respect to it in turn. We then return the set of measurements that are likely given the sampled point with the largest support. As explained in §VII-B, RANSAC is especially sensitive to the likelihood threshold and does not check pairwise consistency.

For PCM, we present results using the exact maximum clique algorithm (PCM-Exact), as well as the heuristic algorithm (PCM-HeuPatt) as explained in §V.

A. Simulated 1D World

First, we simulated a one dimensional world where the robot has a single state variable, x, and receives measurements that are direct observations of that state. We simulate inlier measurements by drawing multiple samples from a Gaussian with a fixed variance and mean x. We simulate both random and perceptually aliased outliers by drawing multiple samples from a single Gaussian with fixed mean and variance and several others from individual Gaussians with random means and variances. We assume the variances are known and are used when computing Mahalanobis distance.

1) Comparison with Combinatorial: For this first experiment, we compare how well PCM-Exact and PCM-HeuPatt approximated the combinatorial gold standard in (5). We generated 100 000 sample worlds. On each of these samples, we estimated the pairwise consistent set using the combinatorial solution as well as PCM-Exact, PCM-HeuPatt, SCGP [16], and RANSAC [10].

Fig. 2 shows a comparison between these four methods with respect to the combinatorial solution. Both PCM method enforce consistency of the returned measurements. PCM-Exact returns the same number of points as the combinatorial solution 100 percent of the time, while PCM-HeuPatt returns the same number of points 98.97 percent of the time. SCGP varies significantly in both the number of points returned and the consistency of those measurements. RANSAC also sometimes returns more or less points than the combinatorial solution and also fails to enforce measurement consistency.

Interestingly, RANSAC is especially dependent on threshold value. The threshold value for RANSAC is centered around a single point and thus is not the same as the threshold value for PCM. If the value is set too high, the number of inconsistent measurements increases. If it is set too low, the total number of returned measurements decreases below the optimal. In Fig. 2, RANSAC's threshold is set arbitrarily to show a single snapshot.

2) Timing Comparison: We also used this 1D World to evaluate the timing characteristics of the different algorithms. To test this, we generated 500 sample worlds each for an increasing number of measurement points. The results are shown in Fig. 3. Note, these timing results can be significantly improved through parallelization.

B. Synthetic 2D Comparison

To test our method's accuracy and consistency on a full SLAM dataset, we took a portion of the City10000 dataset released with iSAM [2] and split it to form two separate robot trajectories. After removing all factors connecting the two graphs, we generated 81 different versions of this dataset by randomly selecting a subset of the true loop closures between

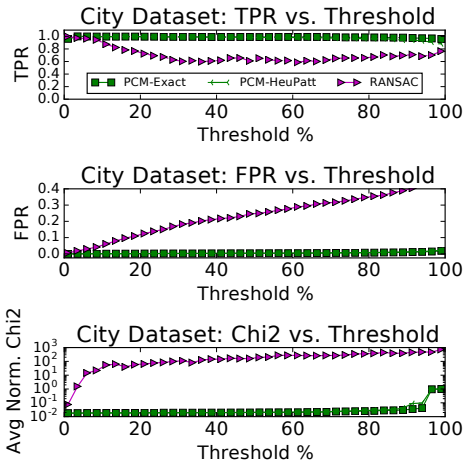

Fig. 4. The true positive rate (TPR = TP / (TP + FN)), false positive rate (FPR = FP / (FP + TN)), and average normalized chi-squared value (Chi2) of PCM-Exact, PCM-Heu, and RANSAC versus the threshold value γ. The TPR and FPR can be thought of as the probability of getting a true positive or a false positive. The Chi2 value should be close to zero if the measurements in the graph are consistent.

the two graphs to be used as inliers, as well as randomly adding outlier loop closures to the graph. As before, some of the outliers are internally consistent to simulate perceptual aliasing and some are generated randomly with random mean and covariance. In this experiment, the number of inlier loop closures was 15, there were two groups of 5 perceptual alliased outliers, and the number of random outliers was 90.

1) Parameter Sweep: Because RANSAC is significantly dependent on the threshold value set, we ran a parameter sweep for the likelihood threshold over all 81 datasets. Fig. 4 summarizes this experiment. The true positive rate (TPR) and false positive rate (FPR) of PCM is relatively unaffected by the choice of the threshold parameter as long as it is less than about 85 percent. RANSAC, on the other hand, has a different FPR for each threshold selected and never has an FPR of zero. This is because PCM conservatively evaluates the consistency of each measurement and determines consistency of a group of measurements as a whole, while RANSAC selects the largest set of measurements that are likely given a single randomly selected measurement. The last plot shows the average normalized chi-squared value of the residual for the entire graph after solving with the selected factors. This value should be close to zero if the graph is consistent.

The results show that PCM does significantly better at restricting the set of measurements to those that are consistent with one another, decreasing the likelihood of getting a false measurement. This is essential because of the extreme susceptibility of SLAM to false loop closures. PCM-HeuPatt is also almost indistinguishable from PCM-Exact.

2) Accuracy Analysis: To evaluate the accuracy of PCM, we compared its performance on all 81 datasets to SCGP, RANSAC (using the two second to lowest thresholds from Fig. 4), and dynamic covariance scaling [6] or DCS. γ for both PCM-Exact and PCM-HeuPatt was set so that it corresponded to the equivalent of 11% likelihood.

Table I gives an overall summary of the results. We used the Mean Squared Error (MSE) of the trajectory of the two graphs (with respect to the no-outlier case), the residual, and the normalized chi squared value of the non-linear least squares solver as metrics to evaluate the solution accuracy. The rotation MSE was calculated via $\epsilon_{rot} = \frac{1}{n}\sum_i \left|\left|log(R_{i\_true}^{\top} R_{i\_est})\right|\right|_F$ over each pose i and the translation MSE was calculated in the normal manner. For this experiment all MSE were calculated with respect to the absolute trajectory value.

PCM has the lowest trajectory MSE, and DCS has the lowest residual. Note that DCS also has the highest trajectory MSE, which is as expected. DCS seeks to minimize the least squares residual error and depends on a good initialization to determine what measurements are consistent enough to not be turned off. Without this initialization, DCS has no reason to believe that the inter-map factors are not outliers and thus turns off all of the inter-map factors in the graph.

RANSAC and both PCM methods take about the same amount of time to find the consistency set once given the matrix $\mathbf{Q}$. The average time to estimate the Mahalanobis distances without the use of analytical jacobians, parallelization, and incremental updates was 70.8 seconds.

Fig. 5 shows example plots of the estimated maps. Both SCGP and RANSAC have trouble disabling all inconsistent measurements. PCM-Heu well approximates PCM-Exact and both PCM methods do well at disabling inconsistent measurements. When PCM does accept measurements not generated from the true distibution, they are still consistent with the uncertainty of the local graphs.

C. Real-World Pose-Graph SLAM

We evaluate PCM on the 3D University of Michigan North Campus Long-Term Vision and LiDAR Dataset (NCLT) [23]. The NCLT dataset was collected using a Seqway robot equipped with a LiDAR and Microstrain IMU, along with a variety of other sensors. There are 27 sessions in all with an average length of 5.5 km per session.

For our experiment, we took two sessions collected about two weeks apart, removed the first third of one and the last third of the other, and then generated potential loop closure measurements between the two graphs by aligning every fourth scan on each graph using GICP [24] and selecting the match with the lowest cost function. We then labeled these registrations as "inliers" and "outliers" by thresholding the translation and rotation mean squared error of the estimated pose transformations with respect to the ground-truth poses for the dataset derived by performing pose-graph optimization on all 27 sessions. Finally, to increase the difficulty of the dataset, we removed all but one sixteenth of the measurements labeled as "inliers" from the graph, resulting in a graph with 10 "inliers" and 98 "outliers."

In this experiment, we compare PCM-HeuPatt with DCS, SCGP, and RANSAC. Fig. 6 shows the normalized chi-squared value of the resulting graphs for RANSAC and PCM versus threshold. Table II provides a comparison of results and Fig. 7 shows the estimated maps. The MSE was calculated using the same method as in the prior section, however in this test we calculated trajectory and map relative pose error seperately. The trajectory MSE calculates the error

TABLE I

RESULTS FROM USING DCS [6], SCGP [16], RANSAC [10](WITH TWO DIFFERENT THRESHOLDS), AND PCM TO ROBUSTLY MERGE MAPS GENERATED FROM A SYNTHETIC CITY DATASET. THESE RESULTS ARE A SUMMARY OF RUNS ON 81 DIFFERENT GENERATED DATASETS. WE EVALUATED THE MEAN SQUARED ERROR (MSE) OF THE TWO GRAPHS WITH RESPECT TO THE NON OUTLIER CASE (NO-OUT). THE WORST RESULTS FOR EACH METRIC ARE SHOWN IN RED, THE BEST ARE SHOWN IN **BLUE**, AND THE SECOND BEST SHOWN IN **BOLD**.

	Trans. MSE (m^2)		Rot. MSE		Residual		Inliers		Chi2 Value		Eval Time (sec)
	Avg	Std	Avg	Std	Avg	Std	TPR	FPR	Avg	Std	Avg
NO-OUT	0.0	0.0	0.0	0.0	32.320	0.117	1.0	0.0	N/A	N/A	N/A
DCS	183077.917	1194931.105	4.169	3.285	**31.687**	**0.507**	0.0	**0.0**	**0.013**	**< 0.001**	N/A
SCGP	623.278	1278.493	0.648	1.535	237385.743	894303.187	0.668	0.051	96.734	364.427	0.006
RANSAC-1%	**5.688**	**21.976**	**0.009**	**0.040**	185.190	587.590	**0.998**	0.006	0.076	0.239	**< 0.001**
RANSAC-3.5%	183.150	636.441	0.236	0.791	3807.570	18478.340	0.974	0.019	1.552	7.530	**< 0.001**
PCM-Exact-11%	**0.276**	**1.537**	**< 0.001**	**0.003**	**45.057**	**105.385**	**0.997**	**0.001**	**0.018**	**0.043**	**< 0.001**
PCM-HeuPatt-11%	**0.276**	**1.537**	**< 0.001**	**0.003**	**45.057**	**105.385**	**0.997**	**0.001**	**0.018**	**0.043**	**< 0.001**

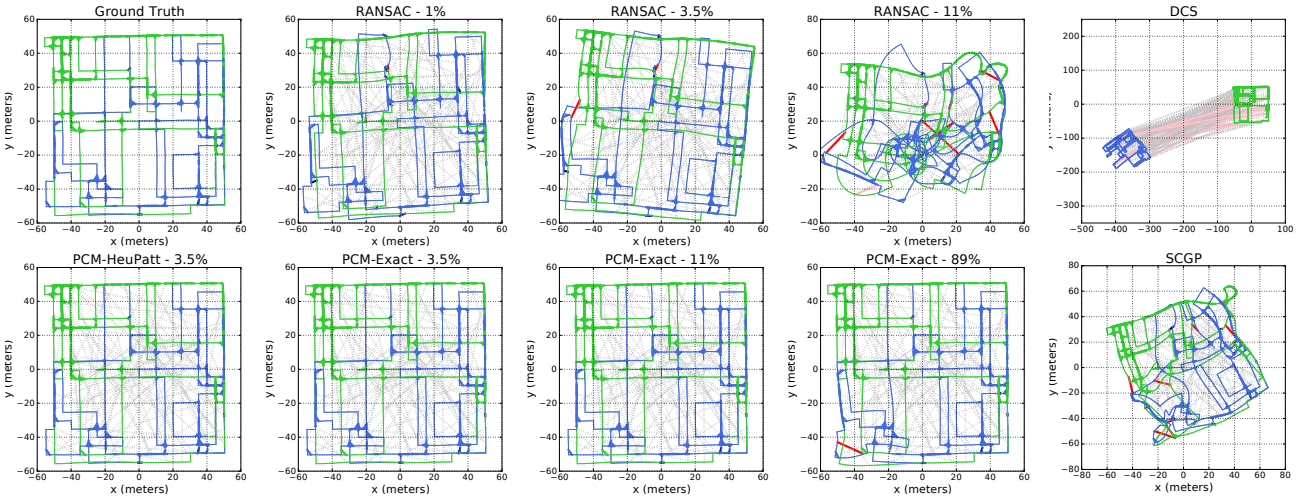

Fig. 5. Example plots of the maps estimated by PCM-Exact, PCM-HeuPatt, RANSAC, DCS, and SCGP for one of the generated city datasets. Correctly labeled inlier factors are shown in bold dark blue with correctly disabled outliers shown as dotted gray. Accepted outliers are shown in bold red with disabled inliers shown in pink.

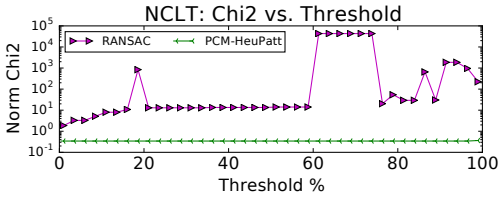

Fig. 6. The normalized chi-squared value (Chi2) of the resulting NCLT graph versus the threshold value γ for PCM-HeuPatt and RANSAC. The Chi2 value for RANSAC is never below one signifying that the selected factors are probabalistically inconsistent, while the Chi2 value for PCM is relatively constant and below 1.0 regardless of threshold.

in the estimated relative pose between consecutive nodes allowing us to evaluate graph correctness, while the relative map pose MSE evaluates the offset between the maps.

PCM results in the graph with the best trajectory MSE and the best translational MSE for the relative pose of the two graphs and results in a consistent graph regardless of threshold. It also detects all of the "inlier" measurements as well as three of the measurements labeled as "outliers". DCS once again disables all measurements. SCGP results in a good graph but only enables three of the "inlier" measurements, and finally RANSAC (for both of the lowest thresholds tried) enables all "inliers" and several "outliers" and results in an inconsistent graph regardless of the threshold selected.

Note that while in this experiment PCM admits more false positives than in the last experiment, the measurements it accepts are consistent with the "inlier" measurements and local trajectories even though they were labeled as "outliers" (Fig. 6). In fact, notice that PCM has a better MSE for the relative map pose then the no outlier (NO-OUT) version of the graph. This suggests that by maximizing the consistent set, PCM is selecting measurements that are actually inliers but were mis-labeled as outliers when compared to the ground-truth. After verification this turned out to be the case.

It is also important to note that although SCGP results in a good graph for this dataset, as shown in the earlier experiment, this does not occur in all cases. In addition, it fails to select the maximum consistent set of measurements, this can be catastrophic in the case of perceptual aliasing.

VIII. CONCLUSION

In this paper, we introduced a novel method called Pairwise Consistency Maximization, or PCM. Using algorithms developed to solve the maximum clique problem, we can quickly estimate the largest consistent set of measurements given a matrix of pairwise consistency measures. This allows us to robustly merge maps generated by multiple robots or multiple mapping sessions without assuming a prior on relative pose. We compare our proposed methods to existing methods such as DCS, SCGP, and RANSAC and show that we significantly outperform these methods.

While in this paper we assume measurements are full degrees of freedom, these methods can be generalized to non-

TABLE II

RESULTS FROM USING DCS [6], SCGP [16], RANSAC [10](WITH TWO DIFFERENT THRESHOLDS), AND PCM TO ROBUSTLY MERGE SEGMENTS EXTRACTED FROM TWO SESSIONS OF THE NCLT DATASET [23]. NO-OUT CORRESPONDS TO A VERSION WITH NONE OF THE MEASUREMENTS LABELED AS OUTLIERS. WE EVALUATED THE MEAN SQUARED ERROR (MSE) OF THE TWO GRAPHS WITH RESPECT TO THE GROUNDTRUTH. THE WORST RESULTS FOR EACH METRIC ARE SHOWN IN RED AND THE BEST ARE SHOWN IN **BLUE**.

	Rel. Pose MSE		Traj. MSE		Residual	Inliers		Chi2	Evaluation
	Trans. (m^2)	Rot.	Trans. (m^2)	Rot.	Error	TP	FP	Value	Time (sec)
NO-OUT	455.4763	0.0308	0.0501	0.0005	765.072	10	0	0.3428	N/A
DCS	206782.2303	0.7154	0.0502	**0.0005**	**724.061**	0	**0**	**0.2568**	N/A
SCGP	522.2352	**0.0162**	0.0502	**0.0005**	748.351	3	**0**	0.3417	0.0021
RANSAC - 1%	1244.3818	0.0697	0.1036	0.0015	4228.21	**10**	6	1.8643	$<$ **0.0001**
RANSAC - 3.5%	13507.7032	17.4156	0.1146	0.0040	7457.54	**10**	7	3.2795	0.0001
PCM-HeuPatt	**386.6876**	0.0245	**0.0501**	**0.0005**	817.803	**10**	3	0.3635	0.0001

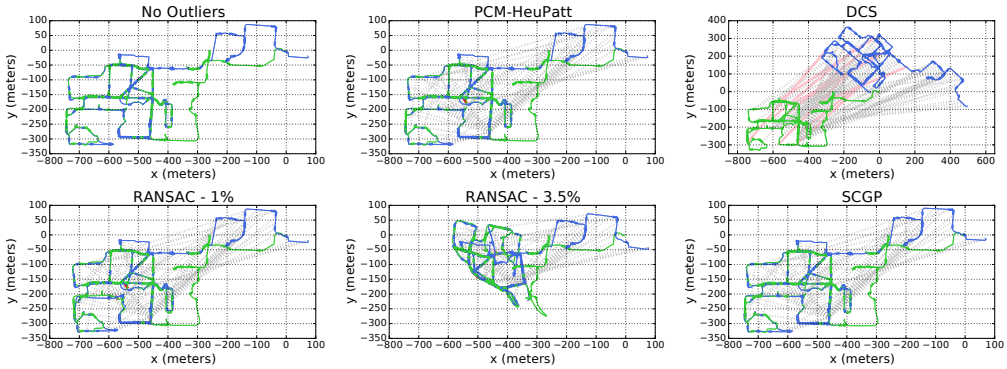

Fig. 7. Plots of the trajectories of two partial sessions of the NCLT dataset as estimated by PCM-HeuPatt, RANSAC, DCS, and SCGP.

full degree of freedom measurements by defining group-k consistency over generalized graphs. In addition, developing methods for using PCM in an incremental fashion would be an interesting avenue for future work.

REFERENCES

[1] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[2] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Trans. on Robotics*, vol. 24, no. 6, pp. 1365–1378, 2008.

[3] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in *Proc. IEEE Int. Conf. Robot. and Automation*, Shanghai, China, May 2011, pp. 3607–3613.

[4] N. Sünderhauf and P. Protzel, "Towards a robust back-end for pose graph SLAM," in *Proc. IEEE Int. Conf. Robot. and Automation*, St. Paul, Minnesota, USA, May 2012, pp. 1254–1261.

[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] P. Agarwal, G. D. Tipaldi, L. Spinello, C. Stachniss, and W. Burgard, "Robust map optimization using dynamic covariance scaling," in *Proc. IEEE Int. Conf. Robot. and Automation*, Karlsruhe, Germany, May 2013, pp. 62–69.

[7] Y. Latif, C. Cadena, and N. José, "Robust loop closing over time," in *Proc. Robot.: Sci. & Syst. Conf.*, Sydney, Australia, July 2012.

[8] M. Pfingsthorn and A. Birk, "Generalized graph SLAM: Solving local and global ambiguities through multimodal and hyperedge constraints," *Int. J. Robot. Res.*, vol. 35, no. 6, pp. 601–630, 2016.

[9] P. Ozog, N. Carlevaris-Bianco, A. Kim, and R. M. Eustice, "Long-term mapping techniques for ship hull inspection and surveillance using an autonomous underwater vehicle," *J. Field Robot.*, vol. 33, no. 3, pp. 265–289, 2016.

[10] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[11] J. Neira and J. D. Tardós, "Data association in stochastic mapping using the joint compatibility test," *IEEE Trans. on Robotics*, vol. 17, no. 6, pp. 890–897, 2001.

[12] 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. Robot. and Automation*, Seattle, Washington, USA, May 2015, pp. 5807–5814.

[13] V. Indelman, E. Nelson, J. Dong, N. Michael, and F. Dellaert, "Incremental distributed inference from arbitrary poses and unknown data association: Using collaborating robots to establish a common reference," *IEEE Control Syst. Mag.*, vol. 36, no. 2, pp. 41–74, 2016.

[14] L. Carlone, A. Censi, and F. Dellaert, "Selecting good measurements via 1 relaxation: A convex approach for robust estimation over graphs," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, Chicago, Illinois, USA, September 2014, pp. 2667–2674.

[15] E. Olson, "Recognizing places using spectrally clustered local matches," *Robot. and Auton. Syst.*, vol. 57, no. 12, pp. 1157–1172, 2009.

[16] E. Olson, M. R. Walter, S. J. Teller, and J. J. Leonard, "Single-cluster spectral graph partitioning for robotics applications." in *Proc. Robot.: Sci. & Syst. Conf.*, Cambridge, Massachusetts, USA, June 2005.

[17] 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. Robot. and Automation*, Anchorage, Alaska, USA, May 2010, pp. 3185–3192.

[18] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," *Auton. Robot.*, pp. 167–193, 1990.

[19] Q. Wu and J.-K. Hao, "A review on algorithms for maximum clique problems," *European J. of Oper. Res.*, vol. 242, no. 3, pp. 693–709, 2015.

[20] D. Zuckerman, "Linear degree extractors and the inapproximability of max clique and chromatic number," in *Proc. ACM Ann. Symp. Theo. Comp.*, Seattle, Washington, USA, May 2006, pp. 681–690.

[21] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy, "Approximating clique is almost np-complete," in *Proc. IEEE Ann. Symp. Found. Comp. Sci.*, Kazimierz Dolny, Poland, Sept. 1991, pp. 2–12.

[22] B. Pattabiraman, M. M. A. Patwary, A. H. Gebremedhin, W. keng Liao, and A. Choudhary, "Fast algorithms for the maximum clique problem on massive graphs with applications to overlapping community detection," *Internet Mathematics*, vol. 11, no. 4-5, pp. 421–448, 2015.

[23] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of Michigan North Campus long-term vision and lidar dataset," *Int. J. Robot. Res.*, vol. 35, no. 9, pp. 1023–1035, 2015.

[24] A. Segal, D. Haehnel, and S. Thrun, "Generalized-icp." in *Proc. Robot.: Sci. & Syst. Conf.*, Seattle, Washington, USA, June 2009.