IEEE Robotics and Automation Letters (RAL) paper presented at the
2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Robust Loop Closure Method for Multi-robot Map Fusion by Integration of Consistency and Data Similarity

Haggi Do[1], Seonghun Hong[2], and Jinwhan Kim[1]

***Abstract*— For an efficient collaboration of multi-robot system during missions, it is essential for the system to create a global map and localize the robots in it. However, the relative poses among robots may be unknown, preventing the system from generating the reference map. In such cases, the necessary information must be inferred through inter-robot loop closures, which are mainly perception-derived measurements obtained when robots observe the same place. However, as perception-derived measurements rely on the similarity of sensor data, different places could be wrongly identified as the same location if they exhibit similar appearances. This phenomenon, called perceptual aliasing, produces inaccurate loop closures that can severely distort the global map. This study presents a robust inter-robot loop closure selection for map fusion that utilizes the degrees of both consistency and data similarity of the loop closures for accurate measurement determination. We define the coalition of these information as the measurement pair score and employ it as weights in the objective function of the combinatorial optimization problem that can be solved as maximum edge weight clique from graph theory. The algorithm is tested on an experimental dataset for performance evaluation and the result is discussed in comparison to a state-of-the-art method.**

I. Introduction

For multi-robot systems to efficiently coordinate and collaborate during missions, it is imperative for the systems to estimate where the robots are relative to one another. One way of achieving this is to establish a global reference frame that all the robots can refer to for navigation. A common reference frame can be obtained from GPS but the signals are not always available. In such cases, robots can execute simultaneous localization and mapping (SLAM) and merge the generated local maps into a global map.

For pose graph SLAM, this process, called map fusion, is achieved through relative pose measurements between the local maps and they can be obtained with direct and indirect observations. A direct observation, as the name suggests, is made when a robot recognizes another robot through its perceptual sensors such as cameras or lidars. While this yields a reliable measurement for the calculation of the relative pose between trajectories, it is not always guaranteed that the robots will come into contact with each other. Therefore, it becomes necessary to consider the second method of obtaining relative pose measurements, which is the indirect observation. An indirect observation can be acquired when a robot observes similar features as the ones another robot has from exploring the same location. Hence, given that the robot trajectories have overlaps, it is highly likely to have indirect observations while there may be no direct observation at all.

However, as indirect observations are perception-derived measurements, relying only on the similarity between the sensor data due to the absence of prior information regarding the location, they are prone to perceptual aliasing. Perceptual aliasing is the phenomenon of wrongly identifying different places as the same place, generating incorrect relative pose measurements, and it may occur in highly structured environment or settings that exhibit repetitive patterns and similar appearances. If these incorrect measurements are used, it could severely distort the map and lead to catastrophic failure in the SLAM system as even just a few incorrect measurements can be detrimental in the optimization process. Accordingly, it becomes crucial to select only the correct measurements in order to fuse the local maps into an accurate global map.

This study proposes a loop closure method for robust map fusion that rejects inaccurate loop closures without any prior knowledge of relative poses between local maps by considering the similarity of the data in the measurement and the consistency between loop closures. By integrating the degrees of such information of the loop closures, we define a novel term, measurement pair score, and attempt to find the loop closure set with both high cardinality and measurement pair score. A combinatorial optimization problem by nature, this can be converted into and solved as a maximum edge weight clique problem. As the weighting of the aforementioned properties are accounted for, the correct set of measurements can be selected even in the presence of many perceptually aliased measurements.

II. Related Work

The previously mentioned indirect observation is also called appearance-based place recognition and many studies have been conducted for its vital role in SLAM and navigation in general. [1] and [2] provide thorough surveys of work related to appearance-based place recognition, many of which try to handle this problem of perceptual aliasing. Yet, it is still challenging to be entirely free of the problem and it often requires a very high threshold as in [3] which may result in much fewer measurements. This issue of incorrect

*This work was supported by the project titled "Development of a core technology and infra technology for the operation of USV with high reliability" funded by the Ministry of Oceans and Fisheries, Korea.

[1]Haggi Do and Jinwhan Kim are with the Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea kevindo@kaist.ac.kr; jinwhan@kaist.ac.kr

[2]Seonghun Hong is with the Department of Mechanical and Automotive Engineering, Keimyung University, Daegu 42601, South Korea sh.hong@kmu.ac.kr

loop closure generation, together with the susceptibility of graph SLAM to incorrect measurements, has motivated research in outlier rejection, and there have been studies that aim to do so through the back-end solver. Generalized graph SLAM [4] attempts to detect outlier measurements by modeling ambiguous measurements as multimodal mixture of Gaussian constraints. Switchable constraints [5], [6] suggests using a new variable to adjust the weights of the loop closures depending on the residual error. If a loop closure constraint leads to a high residual error, its weight is lowered so the solution would be unaffected. Modifying these methods, Dynamic covariance scaling [7] allows covariance to be scaled, rendering the switch to become more flexible. Another method that employs weighting on the loop closure strategy computes the weights by using the Expectation Maximization (EM) algorithm [8]. Other than the use of residual error, one may also utilize the concept of consistency between loop closures for outlier rejection. Realizing, Reversing, Recovering [9] makes clusters of loop closures and checks the consistency of the clusters in order to select a set of consistent measurements based on consensus for robust back-end in pose graph SLAM. However, these studies are not specifically designed for multi-robot applications. Therefore, they require the prior information of the relative pose between the local maps which is unavailable for the cases this paper tackles.

A back-end solver approach [10] in consideration of multi-robot implementation attempted to solve the outlier rejection through convex relaxation. A method called Uncertainty-aware EM [11] was proposed for distributed cooperative mapping. It focused on a real-time application by estimating the initial transformation between the local maps and recursively updating the estimate with online EM approach. Single Cluster Graph Partitioning (SCGP) [12], [13] finds a set of inlier measurements by considering consistency between the loop closures. The work that our research builds on presents a method titled Pairwise Consistency Maximization (PCM) [14]. PCM creates a matrix containing consistency values and makes a binary classification on the pairs of measurements as either being consistent or inconsistent. It then finds the largest set of measurements that are consistent with every other element in the set. DOOR-SLAM [15], a SLAM system that integrates a distributed pose graph optimizer and PCM, has been proposed in the context of distributed system with low communication bandwidth. However, while applicable to multi-robot SLAM, these methods do not take data similarity into account. Consequently, the information from the data association part is not fully utilized for discerning the accuracy of loop closures.

III. Multi-robot SLAM and Pairwise Consistency

Here, we explain pose graph SLAM formulation for a single robot and its expansion to multi-robot SLAM. The estimation of the full trajectory of a robot can be formulated as a *maximum a posteriori* (MAP) estimate as follows:

$$\hat{\mathbf{X}} = \arg\max_{\mathbf{X}} p(\mathbf{X}|\mathbf{Z}). \tag{1}$$

$\mathbf{X}$ and $\mathbf{Z}$ refer to the trajectory and the measurements of the robot. $\mathbf{X}$ is composed of $\mathbf{x}_i$, the pose at time step i, and the measurement set $\mathbf{Z}$ consists of $\mathbf{z}_{ij}$ which links the two poses, $\mathbf{x}_i$ and $\mathbf{x}_j$. As is common in many SLAM literature, the noises can be assumed to follow zero-mean Gaussian distribution allowing us to reformulate (1) as a nonlinear least squares problem that can be solved with solvers such as iSAM [16] and g2o [17].

This single robot formulation can be modified into multi-robot SLAM with the addition of a few concepts. First of all, another class of loop closure measurements is introduced. As opposed to the previous formulation that only had loop closures within a single pose graph, defined as intra-robot measurements, there are as many pose graphs as the number of robots which require inter-robot loop closures. Next, to fuse the maps together into a global reference frame, the relative pose between the local coordinates must be known which can be computed using anchor nodes as suggested in [18]. With these additions, (1) can be modified into

$$\hat{\mathbf{X}}^a, \hat{\mathbf{X}}^b, \hat{\mathbf{T}}_a^g, \hat{\mathbf{T}}_b^g = \underset{\mathbf{X}^a, \mathbf{X}^b, \mathbf{T}_a^g, \mathbf{T}_b^g}{\arg\max}\ p(\mathbf{X}^a, \mathbf{X}^b, \mathbf{T}_a^g, \mathbf{T}_b^g | \mathbf{Z}^a, \mathbf{Z}^b, \mathbf{Z}^{ab}), \tag{2}$$

where $\mathbf{X}^a$ and $\mathbf{X}^b$ are the trajectories of robots a and b, $\mathbf{T}_a^g$ and $\mathbf{T}_b^g$ denote the relative transformations from the robots to the global frame, $\mathbf{Z}^a$ and $\mathbf{Z}^b$ are intra-robot loop closure sets, and $\mathbf{Z}^{ab}$ is the inter-robot loop closure set. The elements in the $\mathbf{Z}^{ab}$ can be written as $\mathbf{z}_{ij}^{ab}$ which represents a relative pose measurement between robot a at time i with robot b at time j.

It has already been mentioned that this paper uses consistency between loop closures to identify the correct measurement set. Pairwise consistency is a quantity that describes whether two measurements are in accordance with the known odometry information which we assume to be reasonably accurate. While a few metrics have been suggested, we employ the one proposed in [13] as follows:

$$c(\mathbf{z}_{jl}^{ab}, \mathbf{z}_{ik}^{ab}) = \exp\left(-\left\| \hat{\mathbf{x}}_{ij}^a \oplus \mathbf{z}_{jl}^{ab} \oplus \hat{\mathbf{x}}_{lk}^b \ominus \mathbf{z}_{ik}^{ab} \right\|_{\Sigma}^2 \right). \tag{3}$$

The notations $\oplus$ and $\ominus$ are used for pose composition and inversion as in [19], and $\hat{\mathbf{x}}_{ij}^a$ indicates the pose estimate from time i to j within robot a frame. This metric is proportional to the probability density of returning to the original pose after being propagated by the odometry and loop closure constraints.

With such definition, a condition that describes a pairwise internally consistent set $\tilde{\mathbf{Z}}$ can be summarized as

$$c(\mathbf{z}_i, \mathbf{z}_j) \geq \gamma, \quad \forall \mathbf{z}_i, \mathbf{z}_j \in \tilde{\mathbf{Z}}, \tag{4}$$

where γ represents the threshold for pairwise consistency. This condition, suggested by [14], requires for $\tilde{\mathbf{Z}}$ to only contain the measurements that have pairwise consistency values greater than the threshold γ among themselves. In addition, as the consistency value is computed using the Mahalanobis distance, a suitable γ can be chosen by considering the χ^2 distribution [13].

IV. Proposed Method

As established, erroneous loop closures may be generated. Therefore, instead of adding the entire set $\mathbf{Z}^{ab}$ into the factor graph, only the accurate measurements need to be accepted. PCM attempts to find such measurement set by using the pairwise internal consistency definition in a binary sense. That is, it only checks whether the consistency value has exceeded the threshold, without considering its extent. However, for greater robustness when there are many uncertain measurements, it becomes necessary to weight the measurements depending on the degree of consistency. In addition, the confidence of the data association can also be used to further this measurement weighting scheme. Depending on the environment, the threshold for data association may vary and, if the environment is unknown, a less strict than appropriate threshold may be chosen. However, it is more likely for the true loop closures to yield higher confidence values than the faulty ones. Taking this into account, we propose a weighted approach that considers the consistency of the loop closure set as well as the data similarity, which is related to the likelihood of the loop closure. This leads to a robust selection of the inlier subset of inter-robot loop closures, details of which will be explained in the subsequent derivation.

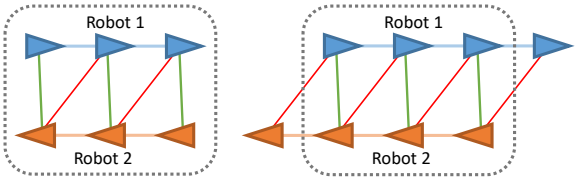
A. Robust Loop Closure Selection Problem

By intuition, one can safely say that accurate loop closures are all consistent with each other. Conversely, it is plausible that a subset of measurements that are not consistent among themselves may contain false measurements. Building on this logic, true loop closure set must be pairwise internally consistent, allowing us to narrow down the search to only such sets rather than considering all the possible combinations.

However, when perceptual aliasing induces copious incorrect loop closures such as in highly structured environments, there may be multiple pairwise internally consistent subsets within the measurement set $\mathbf{Z}^{ab}$. Although we know that the correct solution is one of these subsets, the lack of prior knowledge of the relative offsets between the robots makes it difficult to identify the right one. Accordingly, we attempt to logically distinguish the true subset among many based on certain reasonable assumptions.

Assumption 1: As the inter-robot measurement set size increases, true observations are more likely to accumulate than the erroneous pairwise internally consistent sets.

A similar assumption is made in [14] which implies that the maximum cardinality is the indication of the true pairwise internally consistent set. While it is true that inaccurate loop closures are less likely to be consistent with each other and result in large set size, true measurements may not always have the highest cardinality especially with perceptual aliasing. In cases where the overlap between the maps is sufficiently large, maximum cardinality does indeed correspond to the true set as demonstrated in Fig. 1(a). However, this is not always the case as the shared region between the local maps may be small and this uncertainty is especially exacerbated by the fact that the relative poses are unknown.



(a) Large overlap (b) Small overlap

Fig. 1: Illustration of how degree of overlap may affect the solution's dependence on the consistent measurement set size. The true and false measurements are shown by green and red lines, respectively. The gray box represents the overlapping region.

Hence, due to this uncertainty, measurement set size is not a guaranteed indication of the the true set. For example, in Fig. 1(b), the correct loop closures are outnumbered by the perceptually aliased measurements that have elements from outside the shared environment. Consequently, this urges us to utilize more information.

Given a set of measurements, the true loop closure set must comply with the odometry information more than the false sets, thereby yielding higher pairwise consistency values. Furthermore, every loop closure is a measurement between two instances of data from perceptual sensors. It is made when these data are similar enough that they can be considered to be taken from the same location. Naturally, this implies that data similarity is related to the probability that the loop closure is correct. Hence, among the data associations made, the most reliable measurements would be the ones with the highest degree of data similarity. This confidence of the measurement depends on the type of perceptual sensor and the loop closure detection method used. For a simple example, if a camera is used for indirect observation, the data similarity could be based on the number of feature matches. More feature matches indicate a higher degree of data similarity which means that it is more likely that the images are taken from the same place. In the case of point clouds, low residual error after performing iterative closest point (ICP) would be indicative of high confidence in the loop closure. Considering these aspects, our final assumption is as follows.

Assumption 2: The true measurements are more likely to have higher degree of similarity in the data and consistency values among themselves than perceptually aliased subsets.

Accordingly, we define a new metric, measurement pair score $m(\mathbf{z}_i, \mathbf{z}_j)$, which represents the probability of $\mathbf{z}_i$ and $\mathbf{z}_j$ being both true and consistent. Based on these assumptions and definition, we form the following objective function:

$$\tilde{\mathbf{Z}}^* = \underset{\tilde{\mathbf{Z}} \in \mathbf{Z}^{ab}}{\arg\max} \sum_{\mathbf{z}_i, \mathbf{z}_j \in \tilde{\mathbf{Z}}} m(\mathbf{z}_i, \mathbf{z}_j). \tag{5}$$

As expressed in (5), we search for the pairwise internally consistent set with the highest linear sum of measurement pair score. This is due to the fact that this condition considers the assumptions pertaining to the true loop closure set.

In order to maximize the sum, there must be a lot of elements in the subset, satisfying ***Assumption 1***. Moreover and obviously, a large linear sum is an indication that each measurement pair score is large, which is in accordance with ***Assumption 2***. Simply put, a large linear sum requires both the measurement pair score and the set size to be large.

For the computation of $m(\mathbf{z}_i, \mathbf{z}_j)$, although consistency and measurement confidence are not probabilistically independent, we use the following approximation for the sake of simplicity:

$$m(\mathbf{z}_i, \mathbf{z}_j) = s(\mathbf{z}_i)c(\mathbf{z}_i, \mathbf{z}_j)s(\mathbf{z}_j), \tag{6}$$

where $s(\mathbf{z}_i)$ represents the degree of data similarity of the measurement $\mathbf{z}_i$. As mentioned, pairwise consistency value is proportional to a probability density and data similarity is related to the confidence of the loop closure $\mathbf{z}_i$ based on the perception derived measurement method. While (6) is not exact, it provides the intuition of joint probability and can be easily calculated. Substituting (6) in (5) yields

$$\tilde{\mathbf{Z}}^* = \underset{\tilde{\mathbf{Z}} \in \mathbf{Z}^{ab}}{\arg\max} \sum_{\mathbf{z}_i, \mathbf{z}_j \in \tilde{\mathbf{Z}}} s(\mathbf{z}_i)c(\mathbf{z}_i, \mathbf{z}_j)s(\mathbf{z}_j). \tag{7}$$

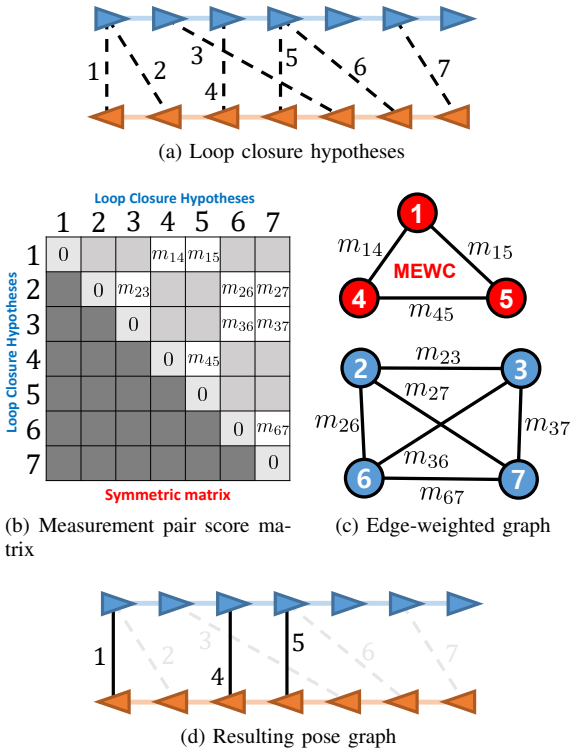
In summary, we defined the best inter-robot loop closure set as the one that satisfies (4) and (7), which is a combinatorial optimization problem that can be solved by leveraging graph theory.

B. Leveraging Graph Theory for Map Fusion

In an undirected graph, edges can be assigned weights denoted as $w(u, v)$ where u and v represent the vertices. Such edge-weighted graph can be expressed using a weighted adjacency matrix in which weight of 0 represents the absence of an edge. Within the graph, a clique Q is defined as a subset of fully connected vertices. Among many cliques that can be found in a graph, the clique that corresponds to the highest sum of edge weight is defined as the maximum edge weight clique (MEWC).

With these concepts of cliques, we discuss how the robust loop closure selection is reformulated into an MEWC problem using a simple example. This subsection focuses on the algorithm explanation of outlier rejection for a pair of pose graphs and the extension to multiple maps is detailed in the subsequent subsection. We assume that there are two relative pose graphs as shown in Fig. 2(a) with 7 inter-robot loop closure hypotheses in the set $\mathbf{Z}^{ab}$. First, for every pair of measurements, pairwise consistency value c is calculated using (3) and these values are stored in a matrix. This matrix is symmetric and the diagonal terms are zero as consistency with itself is trivial and unnecessary. Next, we change all the values in the matrix that are smaller than a predefined threshold into zeroes and define it as the consistency matrix $\mathbf{C}$. Then, we construct another matrix whose diagonal terms hold all the data similarity values of the loop closures and define it as the similarity matrix $\mathbf{S} = diag([s(1), s(2), s(3), s(4), s(5), s(6), s(7)])$. Finally, we apply the following equation:

$$\mathbf{M} = \mathbf{SCS}, \tag{8}$$



(a) Loop closure hypotheses

(b) Measurement pair score matrix

(c) Edge-weighted graph

(d) Resulting pose graph

Fig. 2: Proposed algorithm overview

where $\mathbf{M}$ represents the measurement pair score matrix. This newly constructed matrix is in the form of a weighted adjacency matrix, meaning that there is an undirected edge-weighted graph corresponding to $\mathbf{M}$ that looks like Fig. 2(c).

It can be observed from this graph that any clique corresponds to a pairwise internally consistent set. The analogy is that the loop closures become the vertices in the undirected graph and the existence of an edge between two measurements is contingent on whether the pairwise consistency value exceeds the threshold or not. Due to the fact that clique's underlying definition is the full connectivity, the loop closures in a clique are all pairwise consistent with all of the elements in the subset.

Furthermore, the weights of the edges are the measurement pair scores. Hence, following this analogy, it becomes apparent that the MEWC in the graph gives the subset that meets the objective function (5) because it has the largest linear sum of the measurement pair score values. In addition, as MEWC is a clique, it satisfies the condition of being pairwise internally consistent in (4). Therefore, the best loop closure $\tilde{\mathbf{Z}}^*$ corresponds to the measurements that belong to the MEWC of the edge-weighted graph as follows:

$$\tilde{\mathbf{Z}}^* = \text{MEWC}. \tag{9}$$

Now that the robust loop closure selection has been performed, for map fusion, the accepted inter-robot measurements are added to the factor graph and optimized. If the MEWC corresponds to vertices 1, 4, and 5 for this particular example, then the pose graph should look like Fig. 2(d).

In calculating MEWC, any of the relevant algorithms can be used. Among many algorithms, we utilize MECQ [20] as it is one of the recently proposed methods that extracts the exact solution and provides the source code of the implementation.

C. Expansion to Multiple Pose Graphs

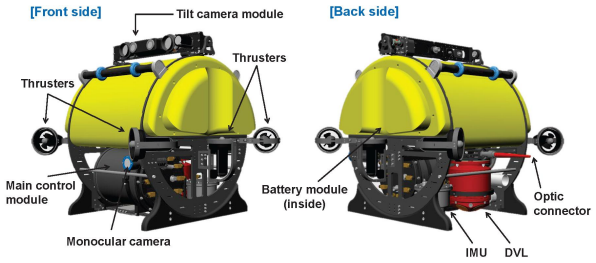
The proposed method of pairwise map fusion can be extended to multiple pose graphs. For every pair of local maps, a consistency matrix $\mathbf{C}$ and a similarity matrix $\mathbf{S}$ can be constructed, and the MEWC can be computed along with the corresponding measurement pair score weight sum. Once all the MEWCs and the weights are found, two local maps with the highest sum of measurement pair score can undergo map fusion first, and the newly formed map can be seen as another local map. This procedure can be iterated until all the local maps are combined into a single global map. It is important that more certain map fusions with higher weight sum are taken place first because subsequent map fusions are more likely to be compromised if maps are incorrectly merged in the beginning. Also, having a larger local map through a successful map fusion increases the chance of increasing the overlap between pairs of local maps. This would lead to greater certainty in the next map fusion iteration, as the cardinality of the pairwise internally consistent set would increase, ultimately enlarging the measurement pair score sum.

V. Performance Evaluation

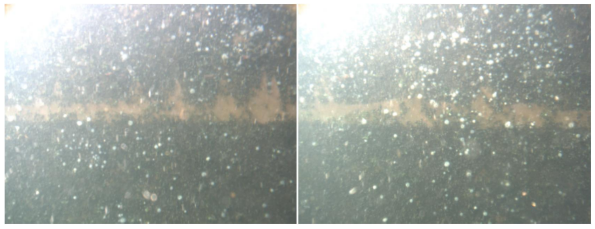
To demonstrate the viability of the proposed algorithm, its performance was assessed using an underwater experiment dataset. In this experiment, an AUV specifically designed for ship hull inspection was deployed and made to follow a lawnmower pattern while keeping a fixed distance from the ship. Fig. 3 shows the actual platform and the images taken during the experiment.

What was challenging about this dataset was the absence of the ground truth in examining the navigational performance. However, as this study focuses on the outlier rejection of inter-robot measurements rather than the navigational performance of each robot, we devised a scenario that allows us to determine with certainty whether a loop closure is flawed or not. The scenario involved dividing a single session's sensor data log to form two segments so that we can emulate a situation that has two AUVs executing the mission. As this single trajectory is used to build two separate trajectories, it becomes relatively easy to determine whether an inter-robot measurement is accurate or not as the relative pose between the pose graphs is known. In addition, to avoid bias from perception, only one side of the stereo images was used for each dataset (i.e. only left images for AUV 1 and only right images for AUV 2).

For inter-robot loop closure hypotheses generation, we extracted features in the image candidates and performed feature matching through RANSAC. While there are many features proposed in the field of computer vision, we employed SURF [22] in this particular implementation as it



(a) Sensors and actuators of the AUV



(b) Images taken from two different places

Fig. 3: AUV and the actual underwater images taken [21]

is often used for underwater applications. With the matched features, the relative poses were inferred in the same manner as detailed in [23]. As for the data similarity, the following equation was used:

$$s(\mathbf{z}) = \exp\left(-\frac{\theta}{r(\mathbf{z})}\right), \tag{10}$$

where $r(\mathbf{z})$ denotes the number of features matched and θ is the tuning parameter for loop closure detection. θ represents the minimum number of feature matches through RANSAC for the measurement to be considered as a loop closure candidate, and it was set to 10 in this study. With the obtained inter-robot loop closure candidates, outlier rejection was performed. Then, using the selected loop closures, the pose graphs were optimized through open source implementations of iSAM [16] and anchor nodes [18].

A. Small overlap in the map

First, we created a case where the overlap between the pose graphs is small relative to the size of the local maps generated. After conducting the inter-robot loop closure extraction process, a total of 232 measurement hypotheses were created, 58 of which were accurate. The fact that only a portion of the hypotheses was true is the evidence of perceptual aliasing occurrence, and it was found that many of these observations were also consistent due to the repetitive pattern on the ship hull.

For a more thorough evaluation, we also implemented RANSAC, SCGP, and PCM to compare the findings. The map fusion results using each of the mentioned algorithms are shown in Fig. 4. The trajectories drawn are the estimated poses of the cameras, and as AUV 1 used the left images while AUV 2 used the right images of the stereo camera, the accurate estimated pose of AUV 2 should be $0.12m$, the baseline of the stereo camera, to the right of AUV

TABLE I: Quantitative analysis for small overlapping case

	TPR	FPR	Comp. time (s)
RANSAC	0.9828	0.1609	0.002
SCGP	0.3621	0.0057	0.005
PCM	0.0	0.3448	0.085
Proposed	1.0	0.0	0.090

1's pose. This is what is meant by ground truth (GT) in Fig. 4, so the black square is where the red square, the beginning of AUV 2 trajectory, should be if the relative pose between the graphs is correctly deduced. However, it appears that none of the methods shows that the black and red squares coincide except for the proposed method, which implies that only the proposed method succeeded in map fusion. In fact, RANSAC and SCGP even caused severe distortion in the local maps. This suggests that some of the chosen measurements did not agree with themselves which is more likely to happen if pairwise internal consistency is not considered. On the other hand, although PCM does not show much visible distortion in the local maps themselves, the relative pose between the trajectories was inaccurate. The cause was the presence of a group of perceptually aliased yet pairwise internally consistent loop closures which was larger than the true set. As PCM views all the measurements equally if they exceed the consistency threshold, cardinality of the pairwise internally consistent set is the sole criteria in the loop closure selection, despite the fact that the true set has higher weighting in both consistency and data similarity with smaller cardinality.

The quantitative result is summarized in Table I for a more comprehensive analysis, and true positive rate (TPR) and false positive rate (FPR) are calculated as below:

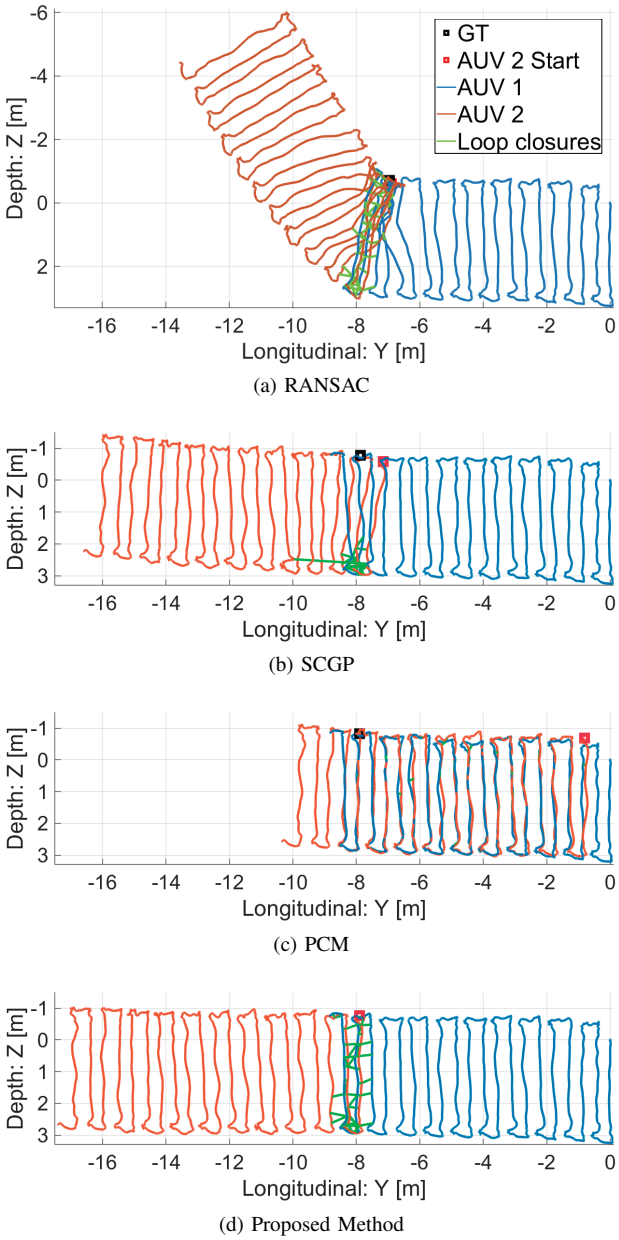
$$TPR = \frac{TP}{TP + FN} \tag{11}$$

$$FPR = \frac{FP}{FP + TN}, \tag{12}$$

where TP, TN, FP, and FN stand for true positive, true negative, false positive, and false negative, respectively. Root Mean Square Error (RMSE) in the relative poses in the overlapping region of the trajectories was found to be $0.0301m$ and 0.2358 for translation and rotation error, respectively. The rotational RMSE was calculated using the following formula:

$$RMSE_{rot} = \sqrt{\frac{1}{n}\sum_{i} \|log(R_i)\|_F}. \tag{13}$$

It can be seen that the proposed algorithm outperforms the other methods by a large margin. Although it took the longest time to compute, it was as fast as PCM and all the true loop closures were selected while rejecting all the incorrect measurements. While one could argue that the proposed algorithm is computationally heavy compared to the others, we would like to emphasize the importance of outlier rejection in map fusion. In the event that a wrong



(a) RANSAC

(b) SCGP

(c) PCM

(d) Proposed Method

Fig. 4: Map fusion result for small overlapping case

set of loop closures is selected, then not only will the map fusion fail but the subsequent true data associations will also be subject to rejection. This is because it is highly plausible that the false loop closures are inconsistent with the accurate ones. Therefore, despite taking longer time for computation, it is more crucial to calculate for the right subset of measurements. Moreover, this algorithm does not have to run every time a new data association is made. Once the algorithm yields the relative pose between the local maps, then, to reject future outliers, it simply has to check if the new hypothesis is pairwise consistent with the already verified measurements.

To also inspect the effect of different similarity values, parameter sweep of θ was executed. As the threshold for

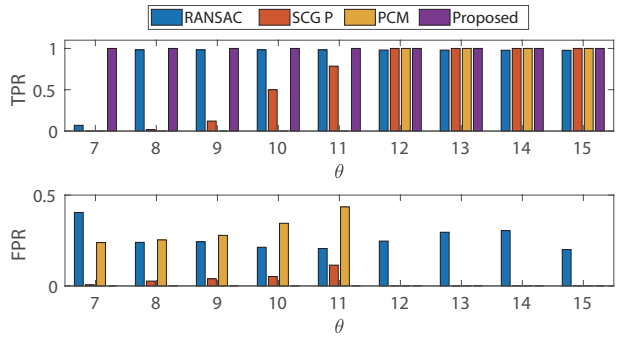
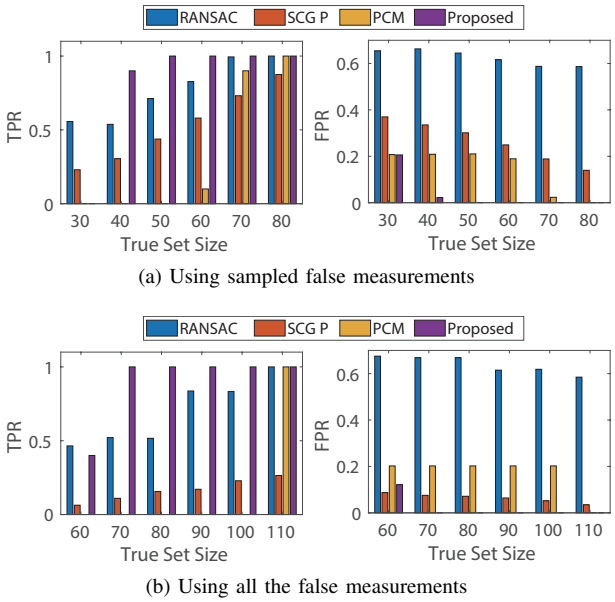


Fig. 5: Outlier rejection performance for different θ values

TABLE II: Results for different θ values

θ	10	11	13	15
Accepted GT ratio (%)	100.00	96.55	86.21	75.86
Trans. RMSE (m)	0.0301	0.0349	0.0368	0.0406



(a) Using sampled false measurements

(b) Using all the false measurements

Fig. 6: Outlier rejection performance comparison

loop closure detection becomes less strict, the number of incorrect measurements increases due to perceptual aliasing. As can be seen in Fig. 5, only the proposed algorithm achieves high performance for a wide range of θ, while the other approaches do so with only stricter thresholds. In the case of SCGP, no measurements are selected when $\theta = 7$ as the high number of false measurements prevents it from having certainty in any of the measurements. Setting an appropriate value of this threshold is not always easy and a high threshold comes at the cost of rejecting true measurements that do not satisfy the stringent criteria for loop closure consideration. Table II summarizes the reduction of accepted true measurements with increased θ where 100% is set as the number of accepted true measurements when $\theta = 10$. Consequently, with fewer true loop closures, RMSE increases.

B. Robustness of the outlier rejection

In order to check the extent of robustness of the proposed algorithm in the presence of perceptually aliased measurements that form internally consistent set with larger cardinality than the true set, more loop closures had to be extracted. Accordingly, we increased the overlap of the map so that it shared half of each trajectory.

With larger overlaps in the map, more loop closures were generated, both true and false. A total of 1010 inter-robot loop closure hypotheses were considered, and it consisted of 491 true measurements and 519 false measurements. It is worth noting that there were multiple large incorrect measurement sets, with maximum size of 105, that were pairwise internally consistent. To see what ratio of true loop closure set size to maximum cardinality caused the MEWC to yield the wrong solution, we varied the ratio of true loop closure to false loop closures and executed outlier rejection.

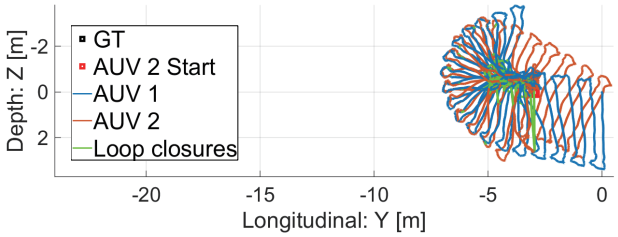
First, we randomly selected 300 incorrect loop closures and various numbers of true measurements, and for each case, a total of 10 different sets were generated. It can be seen in Fig. 6(a) that the proposed method yields much higher TPR and smaller FPR than other methods in most cases. When the ratio of true set cardinality to the largest false set cardinality is as low as 0.6, the proposed method also outputs incorrect solution set. This results in lower TPR than the RANSAC and SCGP, but these methods had significantly higher FPR as well. The reason for failure is that the number of pairwise internally consistent measurements is also considered according to the assumptions established in the previous sections. However, it shows that it is much more robust to the large false measurement set than others, including PCM which is one of the most recent and state-of-the-art methods. Furthermore, the knowledge of the extent of validity in the proposed method can serve as a guide to deciding whether it is safe to conduct map fusion or not.

To further verify that this also applies for larger sets, we ran the same procedure by using all 519 false loop closures found. Similar findings were observed and summarized in Fig. 6(b), and one of the map fusion results is shown in Fig. 7.

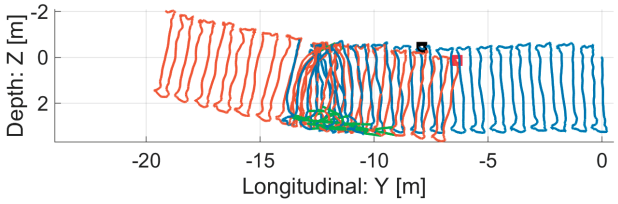
VI. Conclusion

In this paper, a robust inter-robot loop closure selection algorithm for multi-robot pose graph map fusion was proposed. For accurate representation of the environment, the local maps need to be merged into a single reference frame. However, due to certain factors, the necessary information to infer the relative poses between the local maps may be absent which poses as a challenge in the generation of the global map. This issue is exacerbated by the presence of perceptual aliasing and poor data association.

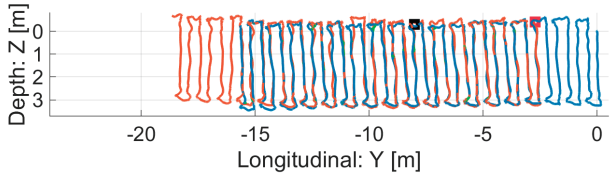
In order to solve this problem, we make the assumption that, among the inter-robot loop closure hypotheses generated, the true measurements form a pairwise internally consistent set with high degree of cardinality, consistency,



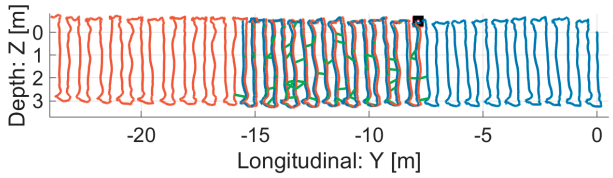
(a) RANSAC



(b) SCGP



(c) PCM



(d) Proposed Method

Fig. 7: Result of map fusion with large overlapping region

and data similarity. This leads us to formulate the problem as a combinatorial optimization with weight consideration of the measurement pair score, a novel definition that is related to the probability that a pair of loop closures is both consistent and true. While this is an exponential operation, we exploit its similarity to a clique problem and show that it can be solved with an MEWC algorithm.

To examine the performance, an underwater field experiment dataset was used and the results of a number of existing works and the proposed method were compared. Both the quantitative and qualitative results of the map fusion showed that the proposed method outperformed the existing algorithms. They also demonstrated that our method was more robust to a much larger number of incorrect measurements that were consistent among themselves.

REFERENCES

[1] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, "Visual place recognition: A survey," *IEEE Transactions on Robotics*, vol. 32, no. 1, pp. 1–19, 2015.

[2] N. Piasco, D. Sidibé, C. Demonceaux, and V. Gouet-Brunet, "A survey on visual-based localization: On the benefit of heterogeneous data," *Pattern Recognition*, vol. 74, pp. 90–109, 2018.

[3] 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," *Journal of Field Robotics*, vol. 33, no. 3, pp. 265–289, 2016.

[4] M. Pfingsthorn and A. Birk, "Generalized graph SLAM: Solving local and global ambiguities through multimodal and hyperedge constraints," *The International Journal of Robotics Research*, vol. 35, no. 6, pp. 601–630, 2016.

[5] N. Sünderhauf and P. Protzel, "Towards a robust back-end for pose graph SLAM," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 1254–1261.

[6] ——, "Switchable constraints for robust pose graph SLAM," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 1879–1884.

[7] P. Agarwal, G. D. Tipaldi, L. Spinello, C. Stachniss, and W. Burgard, "Robust map optimization using dynamic covariance scaling," in *2013 IEEE International Conference on Robotics and Automation*. Citeseer, 2013, pp. 62–69.

[8] G. H. Lee, F. Fraundorfer, and M. Pollefeys, "Robust pose-graph loop-closures with expectation-maximization," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 556–563.

[9] Y. Latif, C. Cadena, and J. Neira, "Robust loop closing over time for pose graph SLAM," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1611–1626, 2013.

[10] L. Carlone and G. C. Calafiore, "Convex relaxations for pose graph optimization with outliers," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1160–1167, 2018.

[11] 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 *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 5807–5814.

[12] E. Olson, M. R. Walter, S. J. Teller, and J. J. Leonard, "Single-cluster spectral graph partitioning for robotics applications." in *Robotics: Science and Systems*, 2005, pp. 265–272.

[13] E. Olson, "Recognizing places using spectrally clustered local matches," *Robotics and Autonomous Systems*, vol. 57, no. 12, pp. 1157–1172, 2009.

[14] J. G. Mangelson, D. Dominic, R. M. Eustice, and R. Vasudevan, "Pairwise consistent measurement set maximization for robust multi-robot map merging," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2916–2923.

[15] P.-Y. Lajoie, B. Ramtoula, Y. Chang, L. Carlone, and G. Beltrame, "DOOR-SLAM: Distributed, online, and outlier resilient SLAM for robotic teams," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1656–1663, 2020.

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

[17] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G 2 o: A general framework for graph optimization," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 3607–3613.

[18] B. Kim, M. Kaess, L. Fletcher, J. Leonard, A. Bachrach, N. Roy, and S. Teller, "Multiple relative pose graphs for robust cooperative mapping," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 3185–3192.

[19] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Autonomous robot vehicles*. Springer, 1990, pp. 167–193.

[20] S. Shimizu, K. Yamaguchi, and S. Masuda, "A maximum edge-weight clique extraction algorithm based on branch-and-bound," *arXiv preprint arXiv:1810.10258*, 2018.

[21] S. Hong, D. Chung, J. Kim, Y. Kim, A. Kim, and H. K. Yoon, "In-water visual ship hull inspection using a hover-capable underwater vehicle with stereo vision," *Journal of Field Robotics*, vol. 36, no. 3, pp. 531–546, 2019.

[22] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[23] S. Hong and J. Kim, "Selective image registration for efficient visual SLAM on planar surface structures in underwater environment," *Autonomous Robots*, vol. 43, no. 7, pp. 1665–1679, 2019.