Robust Distributed 3D Mapping With Communication Constraints

Vibhav Nagaraj Ganesh June 20, 2017 CMU-RI-TR-17-32



The Robotics Institute School of Computer Science Carnegie Mellon University Pittsburgh, Pennsylvania

Thesis Committee: Nathan Michael, *Chair* Artur W. Dubrawski Ben Eckart

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Robotics.

Abstract

Rapid autonomous exploration of challenging, GPS-denied environments such as underground mines provides essential information to search and rescue as well as defense operations. We pursue a distributed perception strategy for a team of robots to develop a consistent distributed map of these communication-constrained environments that in addition exhibit perceptual aliasing due to repetitive structure. Since each robot operates with respect to a local reference frame, robots estimate relative transforms to other robots' reference frames using observed environment correspondences in order to construct a consistent distributed map. Perceptual aliasing, or the incorrect association of observations from different areas in the environment, complicates the estimation of relative transforms between robots. In this work, we extend a robust distributed mapping formulation to operate using 3D sensors under hardware network limitations for operation in the communication constrained, repetitive structure of an underground mine.

Real-world communication constraints limit a robot from sharing large numbers of observations at high fidelity. Naïvely simplifying sensor information leads to loss of unique features and an increase in perceptual aliasing. Towards sharing the most relevant subset of information, we develop a scan utility function based on information theoretic measures to assess a scan's ability to reduce map uncertainty and featurebased place recognition approaches to assess a scan's potential for containing shared observations between robots. Using the utility function to rank scans, we formulate an offer-response-request framework, Communication Constrained Information Routing (CCIR), that ensures operation under stringent bandwidth restrictions. In simulation results, CCIR decreases the required network usage for distributed mapping to 20.7% of a fixed-rate down-selection approach.

Given the ability to share rich 3D information over constrained networks, we pursue full 3D mapping via extensions to existing approaches including robustification techniques. The robust measures we introduce allow operation in the targeted mine environment even given substantial perceptual aliasing with outliers accounting for 98.1% of all detections. Furthermore, the developed CCIR framework allows robots to develop relative transforms while respecting network bandwidth constraints. Similar performance when operating using a fixed-rate down-selection approach over the same mine environment requires 7.69 times more data transmission.

Additionally, to enable operation in environments that exhibit perceptual aliasing that exceeds the performance characteristics of the developed CCIR framework, this thesis details first results for an approach that moves away from feature-based techniques and introduces a methodology utilizing Hierarchical Gaussian Mixture Models. Through regeneration of the point cloud from the HGMM model and Generalized Iterative Closest Point algorithms, we show that we are able to detect multi-robot loop closures accurately with an outlier rate 34% of that of feature-based methods.

Acknowledgements

First thanks to Nate for providing me with research interest and direction. I've been working with him for 2.5 years and have visibly grown throughout my time in his lab.

Thank you to all my wonderful labmates who have worked alongside me. Specifically I wish to acknowledge Cormac O'Meadhra, Wennie Tabib, and Shobhit Srivastava for their direct contributions towards my thesis. Furthermore, I would like to thank Ellen Cappo, Vishnu Desaraju, and Xuning Yang for reviewing this thesis document.

Thanks to all my friends both close and afar who have helped motivate me when I needed it.

Contents

1	Introduction		
	1.1	Previous Work	6
		1.1.1 Distributed Mapping	6
		1.1.2 Network Constrained Systems	7
		1.1.3 Alternative Environmental Models	10
	1.2	Thesis Problem	11
	1.3	Contribution and Outline	12
2	Dis	tributed Mapping Background	14
	2.1	Problem Formulation	15
		2.1.1 Expectation Maximization Formulation for	
		Relative Transform Estimation	17
	2.2	System Details and Naïve 3D Extension	20
	2.3	Results	22
3	Cor	nmunication Constrained Information Routing	24
	3.1	Overview	26
	3.2	Communication Constrained Information	
		Routing	28

		3.2.1	Measurement Evaluation	28
		3.2.2	Evaluating Loop Closure Potential	30
		3.2.3	Evaluating Map Information Potential	31
		3.2.4	Offer-request Paradigm	33
	3.3	Result	ts	34
		3.3.1	Implementation Details and Complexity Analysis	35
		3.3.2	Comparison with Fixed-Rate Down-selection	36
		3.3.3	Simulation Results	38
		3.3.4	Field Experiments	41
	3.4	Chapt	er Summary	43
4	Ext	ending	y the Distributed Manning Framework to 3D	45
-				10
	4.1	3D Fe	atures	46
	4.2	Fast-A	Appearance Based Matching	50
	4.3	Switch	hable Factors	51
	4.4	Result	ts: Mine Environment	53
	4.5	Chapt	er Summary	54
5	Cor	nclusio	n	57
_	5 1	Summ		57
	5.1	Summ		01
	5.2	Future	e Work	60
6	App	pendix		62
	6.1	HGM	M Models	62
	6.2	Novel	Loop Closure Detection	63

List of Tables

3.1	Comparison of CCIR and fixed-rate down-selection over various num-	
	bers of robots	41
4.1	Comparison of FAB Matching and Histogram approach over three mine	
	data sets.	51
6.1	Comparison of HGMM-based approach and FAB Matching	65

List of Figures

1.1	The blue and red robots (triangles) are at different locations in the	
	building yet observe the same local sensor reading (transparent crescent	
	shape). This ambiguity in place recognition is known as perceptual	
	aliasing.	2
1.2	A 3D map of an underground mine. Note the repetitive structure of	
	the columns. Each column is roughly 10m in width	4
1.3	A slice of a 3D Velodyne-16 scan from full resolution through progres-	
	sive down-sampling. Simplifying the scan by down-sampling reduces	
	the texture and uniqueness of the scan. \ldots \ldots \ldots \ldots \ldots \ldots	8
1.4	The entropy of each scan is irregular as the robot operates through the	
	environment.	11
2.1	A potential relative transform between world frame's $(T_{r_1}^{r_2})$ is estimated	
	by a scan correspondence (yellow) between robot r_1 at position $x_k^{r_1}$	
	(blue) and robot r_2 at position $x_l^{r_2}$ (purple)	17

2.2	Each line represents a 2D relative transform where the position of the	
	arrow represents the delta translation and the heading represents the	
	delta yaw. The relative transforms estimated by inlier scan matches	
	are clustered (black circle) while outlier scan matches are randomly	
	distributed.	19

- 2.3 The histogram of matched features over scans identifies potential scan matches [5]. The query scan (blue) matches closely with the local scan (red) at the histogram peak index indicated by the pink arrow. . . . 21
- 2.4 The estimated trajectory of the 3 robots (red, green, blue) with the joint map built by the green robot. This run assumes full communication amongst the agents.23
- 3.2 The different environments used for simulation, including the 2.5D extrusions of (a) a building and (b) a cave where Velodyne-16 scans are simulated as well as a 3D environment with (c) clutter.36

- 3.4 The entropy reduced over time (a) for each approach is comparable while the cumulative network utilization (b) for CCIR is 59% of the that of fixed-rate down-selection. The dark line represents the average while the shaded region represents three standard deviations. 39
- 3.5 Results from two different runs of three different trajectories through the more complicated environment. The figure contains two distributed maps (a), (d) as well as the mean and variance in reduction of entropy over time (b), (e) and the network utilization (c), (f). The colored arrows represent locations where scans are taken. Red represents local 2D scans, yellow is local 3D scans, dark blue is remote 2D scans, and cyan arrows are foreign 3D scans.
 39
- 3.6 Comparison of CCIR and fixed-rate down-selection with six robots in cluttered environment. The figure includes (a) the entropy reduction over time and (b) the network usage as well as (c),(d) the generated maps.
 40

3.7	(a) The entropy reduction over time for a team of 9 robots operating	
	with a Velodyne is comparatively the same between the two approaches	
	while the network utilization (b) for CCIR is 39.5% of the network	
	utilization of fixed-rate down-selection. The required bandwidth was	
	increased to 200 KB/s to allow convergence of the relative transform	
	calculation. \ldots	41
3.8	The robot used for hardware tests is equipped with a bullet wireless	
	transmitter, a Hokuyo 2D Lidar, an Orbecc Astra RGB-D camera and	
	a core i7 Gigabyte Brix for compute.	42
3.9	(a) Map constructed from two ground robots using 2D Lidar and depth	
	camera. (b) Network usage stays below set bandwidth requirements	
	while fixed-rate down-selection requires additional bandwidth to con-	
	struct an equivalent map. \ldots	43
4.1	A normal for a point is the tangent vector of the surface at the point.	
	The blue circle denotes the support of the normal, i.e. the points it	
	uses to estimate the Gaussian which it uses for point-normal calculation.	47
4.2	The unfiltered pointcloud side view (a) and top view (b). After remov-	
	ing normals that have a normalized z-value not within ± 0.5 , we obtain	
	a set of sensible normals, shown in side view (c) and top view (d) $\ . \ .$	48
43	We show the feature correspondences between two scans with (a) and	
1.0	-	
1.0	without (b) normals that are degenerated due to sparsity of the points.	
1.0	without (b) normals that are degenerated due to sparsity of the points. Yellow lines connect points with similar descriptors. While the scan	

4.4	Subfigures depict recognized features (green) in the same environment	
	with successive downsampling: the full scan (a) along with voxel grid	
	down-sampled scans using voxel sizes of 0.1 (b) and 0.2 (c). Downsam-	
	pling reduces the ability of feature detectors to reliably detect interest	
	points if the filter size is large	49
4.5	Histogram of shared features over a robot's scan data base for a query	
	scan (blue). The black bar represents the number of shared features	
	between the query scan (blue) and the scan at the index given by x .	
	The green bars represent scans that are within 3m and are considered	
	ground truth matches. Note how similar false matches look. \ldots .	50
4.6	The pose graph pictured, reproduced from [1], contains switch variable	
	$s_{2,i}$ and switch factor (yellow) that can robustly determine inliers and	
	outliers.	52
4.7	A low prior strength (a) causes all loop closures to be ignored while	
	a high strength prior (c) causes the false positives to be considered as	
	well. A prior tuned for this environment provides an optimized map	
	(b)	53
4.8	The generated map (a), the entropy levels over time (b), and the result-	
	ing network usage (c) from a two-robot trial in the mine environment.	
		54
4.9	The map generated from the fixed-rate down-selection approach (a)	
	and CCIR (b). The two maps are nearly identical; CCIR is able to	
	generate a relative transform using 13.6% of the bandwidth used by	
	fixed-rate down-selection (c).	55

Chapter 1

Introduction

Autonomous agents are becoming increasingly useful at performing surveillance tasks or retrieving information in challenging and hazardous environments such as underground tunnels or disaster areas. In many cases, these environments are unknown apriori and maps must be generated on-line in order for robots to operate and interact with their surroundings. While a single agent is capable of performing mapping operations, hardware issues such as battery capacity and speed restrictions limit its performance. On the other hand, a multi-robot strategy is able to explore more quickly or cover larger areas through effective coordination and sharing in order to build a global model. Furthermore, multi-robot teams are robust to individual robot failures. However, the benefits of a multi-robot system are undercut by the increased computational and algorithmic complexity of having to communicate and coordinate amongst the individual agents. A team of robots requires a shared map and global state information to avoid choosing redundant paths in the context of exploration or for planning coordinated motion. Robots are able to develop a consistent representation of the environment by transmitting local observations and utilizing the shared information from other robots to estimate the global state. Ambiguities in local-



Figure 1.1: The blue and red robots (triangles) are at different locations in the building yet observe the same local sensor reading (transparent crescent shape). This ambiguity in place recognition is known as perceptual aliasing.

ization due to repetitive nature and minimal texture further complicate distributed mapping. While sharing higher fidelity information could help to reduce ambiguities in place recognition, hardware bandwidth frequently precludes robots from sharing all available data. This thesis seeks to develop a robust framework that enables intelligent sharing of information and improves the robustness of distributed mapping to generate rich consistent 3D maps in these challenging environments.

Our proposed approach builds on recent work that utilizes detected correspondences between sensor scans across a pair of robots to estimate a relative transform between each robot's local reference frame. Since robots generate and transmit scans with respect to their local frames of reference, relative transforms are necessary to utilize shared scan information in distributed maps. Methods for global localization like GPS are unavailable in some of the environments of interest in this work, such as underground tunnels, and therefore each robot can only rely on overlapping sensor data between local and received scan information for relative transform calculation. While some approaches rely on rendezvous or robots directly observing other robots, we wish to avoid assumptions requiring the robots to operate in the same area at the same time. Common observations between robots, regardless of at what time sensor measurements were made, allow them to develop constraints that describe their relative trajectories. The recognition of a common observation between robots is termed a multi-robot loop closure, and recognized loop closures allow our team of robots to develop a consistent distributed map. We define a distributed map as the environmental model generated independently on each robot from local and received sensor observations. Each robot constructs a distributed map in its own local reference frame. Therefore, a key requirement of our approach is to generate a set of relative inter-robot transforms that enable robots to interpret measurements collected in another robot's local reference frame and produce a globally consistent distributed map.

However, environments with repetitive structure complicate the task of finding scan correspondences as scans observed from different areas can be mistakenly associated, an issue known as perceptual aliasing (Fig. 1.1). When robots associate sensor observations from different areas of the map, they can generate incorrect relative transform estimates and cause the distributed map to become inconsistent. The joint mapping foundation that we build upon is robust to some number of outlier scan correspondences and enables operation in indoor and feature-rich environments.

Environments such as the mine shown in Fig. 1.2 require additional methodologies due to an increased level of perceptual aliasing. Distinctions between observations taken at significantly different areas in the mine only become apparent when using a high fidelity perceptual model. Therefore, one method of handling the minimally textured environment involves sharing higher resolution information between robots.



Figure 1.2: A 3D map of an underground mine. Note the repetitive structure of the columns. Each column is roughly 10m in width.

However, without existing communication infrastructure, robots must use constrained wireless ad-hoc networks that restrict the sharing of all locally collected sensor information. Therefore, we pursue a strategy to intelligently down-select the information to be shared based on its utility to the developed environmental model. The utility is designed to minimize redundancy in the communicated information while ensuring reliable relative transform calculation. By leveraging the distributed mapping framework developed in [5] and local scan evaluation calculations, the developed framework, called Communication Constrained Information Routing (CCIR) (Sec. 3.2), limits the sharing of sensor observations to respect hardware limitations while preserving the quality of the generated map.

The effectiveness of the communication protocol and scan correspondence detection are highly dependent on the quality of features detected in the environment. Features allow for efficient description of a scan and can be used to quickly evaluate potential overlap in sensor readings. While 2D features are effective for ground robots or in environments with minimal 3D texture, 3D features are necessary in cluttered environments when robots potentially operate at different heights. Chapter 4 introduces the 3D features used for the mine environment pictured in Fig. 1.2. Due to sparsity in the sensor observations as seen in Sec. 4.1, 3D features are inherently noisier and require additional methods to reduce outlier correspondences. Chapter 4 introduces robust measures as described in Sec 4.2 and Sec 4.3 to reduce the implications of outlier correspondences in the mine environment and enable operation given an outlier ratio of 98.1%.

In some cases, however, the outlier ratio is higher than 99.8 % and feature-based approaches are unable to reliably detect loop closures at an accuracy sufficient for relative transform calculation or for CCIR to operate effectively. Towards enabling reliable detection in these scenarios, we explore an alternative map representation strategy based on Hierarchical Gaussian Mixture Models (HGMM) as discussed in Section 6.1. By utilizing a bottom-up Expectation Maximization (EM) approach and simplifying the model construction based on divergence measures, robots are able to develop high fidelity representations of the environment with substantially less memory footprint. Resampling from the HGMM allows us to recreate the original point cloud at a sufficiently high fidelity for effective scan registration with a significantly higher precision than sending quantized versions of the same point cloud. Section 6.2 demonstrates that this loop-closure detection scheme has the potential to produce an order of magnitude of fewer false matches when compared to feature-based approaches for a real-world mine data set.

1.1 Previous Work

1.1.1 Distributed Mapping

The ability of robots to localize with respect to each other and develop a consistent representation of the environment is necessary for effective multi-robot coordination and cooperation in GPS-denied environments. Early works involved with distributed mapping assumed given initial transforms and focused their approach on distributed localization. Pose estimates during online operation are inherently noisy and require an optimization technique to obtain accurate estimates. Towards obtaining an accurate estimate of the global state, initial works began by modifying Extended Kalman Filters [18] or particle filters [10] applied to the distributed setting. However, filter-based approaches do not handle loop closure constraints and cannot robustly determine relative transforms when they are initially unknown. As robot localization drifts over time due to error accumulation, identifying the return to a previously explored location is essential to obtain the most likely estimate of a robot's trajectory. Full-SLAM and Pose-SLAM are recent pose-graph approaches used to account for odometry drift over time by considering all known constraints in an optimization function based on the residuals at each pose. While Full-SLAM optimizes for the landmarks in the environment as well as the poses of the robot, Pose-SLAM only estimates the poses of the robot and is used here to be generic to sensor type and operate in environments without clear landmarks. The approach presented by Kim et al. [15] looks at using multiple relative pose graphs with anchor points to extend the pose-graph to the multi-robot case. This approach however estimates relative transforms from direct observations and is not robust to outlier correspondences and is therefore not applicable to the repetitive structure target environments considered here. To avoid assumptions on a robot's ability to perceive another robot, the proposed approach generates inter-robot constraints indirectly from matching sensor observations. An overview of the distributed mapping work we build upon [5] is presented in Chapter 2.

1.1.2 Network Constrained Systems

In this work, we seek to establish a distributed mapping framework that scales to the available communication capabilities by sharing only measurements relevant to the calculation of relative transformations between robots and construction of the distributed map. As described earlier, the available network hardware bandwidth precludes robots from sharing all local sensor information. To operate within bandwidth limits, robots perform additional computation to reduce the redundancy in transmitted information and share only what is essential. An initial approach seeks to simplify each scan in order to compress the message. Martins et al. [20] present a comparison of compression techniques applied to multi-robot mapping scenarios. However, simplifying a sensor observation through quantization or downsampling in this scenario significantly degrades the multi-robot loop closure detection since perceptual aliasing becomes more of a concern as seen in Fig. 1.3.

One alternative approach to down-sampling is data fusion. Data fusion attempts to reduce shared information without losing perceptual detail by compressing only redundant information. Local maps such as voxel grids already accumulate several measurements and reduce redundancy in storage as points from multiple scans that reside in the same voxel are all represented by that one voxel. Nettleton et al. [22] employ information-theoretic approaches to send subsections of a map between robots and show favorable performance when trajectories have minimal drift and robots have sufficient communication bandwidth. However, in the case of drift in odometry, the



Figure 1.3: A slice of a 3D Velodyne-16 scan from full resolution through progressive down-sampling. Simplifying the scan by down-sampling reduces the texture and uniqueness of the scan.

local submaps will be noisy and lead to incorrect associations as submaps of the same area for two systems will be notably different. Furthermore, while sharing fused measurements can deal with intermittent communication [9] or can be used to reduce transmission frequency to a constant rate [19], constant rate communication does not make optimal use of the available network capacity and cannot respond to dynamic network topologies. Even when local odometry is sufficiently accurate, the sensors considered in this work (specifically 3D Lidar and depth cameras) produce dense measurements that are difficult to transmit. One large observation must be made into smaller packets when transmitted between systems and in scenarios where packet loss occurs frequently, large messages are easily corrupted. Another method for compression given environments with known objects is to use object based models [3]. By transmitting labels rather than the raw sensor data, they show that the required communication is reduced by six orders of magnitude. However, this approach is not amenable in our environments, such as underground tunnels, since the environment does not contain objects that can be learned *a priori*.

Since measurements cannot be heavily quantized or fused, we pursue a strategy that shares a well-chosen subset of all observations. Naïve approaches such as constant rate down-sampling [19] exhibit the same inefficiencies as compression approaches by discarding potentially useful information. Generating accurate relative transforms requires a sufficient number of correct scan correspondences; not directly accounting for potential correspondences decreases the likelihood that a sufficient number of matches needed for the relative transform calculation will be found. In the context of exploration, sensor observations reduce uncertainty of the map and so the degree to which a scan reduces the uncertainty can be used as a measure for scan value. We can see from Fig 1.4 that the value of scans can behave irregularly thereby making fixed-rate down-selection a poor choice. Furthermore, we assume that robots operate in a bounded environment with overlapping paths that result in redundant sensor information similar to the work of Kretzschmar et al [16]. In this work, the authors compress a single robot's pose-graph by pruning scans that insufficiently reduce uncertainty in the map. This approach is readily extended to the multi-robot scenario where overlapping coverage is far more prevalent. In the context of multiple robots, the mutual information of the scan with respect to a robot's local version of the global map can be used to determine which scans are sent across the network or incorporated into each robot's representation of the global map. Paull et al. employ a similar pose-graph pruning approach for multiple Autonomous Underwater Vehicles (AUVs) operating under severely constrained communications [25]. However, the local evaluation of a scan alone does not consider the effect the scan will have on other robots or the current network utilization.

One potential extension for intelligently determining the selection of information to share that takes into consideration the full communications network is to construct a network flow optimization problem where the sources are the sensors and the sinks are the receiving robots [14]. Link costs combine the available bandwidth and estimated utility of a scan. While this approach provides the optimal flow of information given an accurate scan utility function, optimizations across distributed networks suffer from slow convergence times due to latency and changing topologies. We therefore approximate the distributed optimization of information flow through an offer-response-request framework detailed in Section 3.2. In addition, we extend the scan utility function to include the implication of the scan on the relative transform calculation. While our framework does not solve the full network flow problem, it ensures that the network is not over-saturated and allows for real-time performance on existing network hardware by reducing the communication frequency. While this approach works for a wide range of depth sensors, 3D Lidar and depth sensors are utilized in this effort to illustrate the applicability of the methodology.

1.1.3 Alternative Environmental Models

The reliability of the communication approach and the distributed mapping backend require efficient and robust scan correspondence detection. As is shown in Chapter 4, feature-based approaches have fundamental limitations that restrict their functionality in minimally textured, repetitive environments such as our target mine environment. We therefore pursue alternative strategies for map information representation to enable accurate loop closure detection. While it is possible to store map information in a more efficient format than a standard voxel grid using methods such as an Octree, the compression level is minimal and the discretization does not always



Figure 1.4: The entropy of each scan is irregular as the robot operates through the environment.

allow minute details to be captured. An extension called NDT-MAP [7] that places a Gaussian in each cell can be used to add detail to the map at the cost of additional data storage. While each cell is represented by a continuous form, each voxel is still considered independent from other voxels which is not always an accurate assumption. Removing the per-voxel restriction of each Gaussian, we can instead apply an HGMM model [28] that scales to the fidelity required in the environment using information-theoretic approaches. The HGMM model provides a way to compress scan information by two orders of magnitude while preserving the representative power required to identify unique locations within the environment.

1.2 Thesis Problem

This thesis seeks to enable multi-robot mapping in environments with high perceptual aliasing while operating over constrained wireless networks. In order to accomplish this task, we

- develop an information-theoretic system to control the transmission of sensor observations to meet bandwidth requirements and minimize the uncertainty in the localization of the robots and the generated global map, and
- develop a system capable of efficiently and reliably detecting the unique features in the environment in order to precisely associate sensor observations and reduce the effect of outlier scan correspondences.

1.3 Contribution and Outline

Chapter 2 provides a background on the underlying distributed mapping framework. We begin by discussing the problem formulation and the methodology used to allow robust distributed mapping. The naïve extension extracts 2D slices from the 3D scans and allows for initial results which we present in this chapter.

To extend the mapping framework to fully support 3D data, the increased datarate of the sensors must be addressed. Naïve down-selection or down-sampling leads to severely degraded performance. Chapter 3 presents an intelligent approach to limit scan sharing based on information theory along with a comparison to naïve down-selection. Section 3.2.1 describes the development of a measurement utility function through the combination of loop closure potential (Sect. 3.2.2) and novel map information (Sect. 3.2.3). This utility function is used in Section 3.2.4 to evaluate local scans and determine if and how they should be communicated to other robots.

Furthermore, the extension of the distributed mapping front end to 3D relies on new feature and scan-registration algorithms described in Sect. 4.1. In order to reliably map and localize in areas with high perceptual aliasing, additional methodologies are introduced in Sects. 4.2 and 4.3 to improve robustness. Employing these additions, CCIR is applied to the mine environment in Sect. 4.4. and while the overall communication required for reliable mapping is significantly reduced, it is still not supported by current hardware. We conclude in Chapter 5 with the current capabilities of the system and provide initial results of an HGMM mapping framework that mitigates the perceptual aliasing seen in the underground tunnel environment while using significantly less bandwidth.

Chapter 2

Distributed Mapping Background

Distributed mapping seeks to develop a shared environmental model through the communication of local sensor observations by each robot. In order to utilize remote measurements, a local robot requires the location at which the remote measurement was taken. Since robots express their measurements in their local frame, a relative transform between robots is necessary to calculate the remote sensor location in the local frame. We assume that relative transforms are not given a priori and must be estimated online through shared measurements between each pair of robots. A brief introduction to the problem structure is provided in Section 2.1. Section 2.2 details how robots maintain a history of local observations in order to detect overlapping sensor measurements from received messages and construct potential data associations. Due to perceptual aliasing, sensor observations from different areas of the map may be incorrectly associated leading to outlier relative transform estimates. Thus, to accurately construct the true relative transform estimate, multiple consistent associations must be formed along the robots trajectory. We utilize an Expectation Maximization (EM) approach as described in Section 2.1.1 to efficiently and accurately determine the true relative transform.

In Simultaneous Localization and Mapping (SLAM) problems, it is common to separate the problem into two distinct components: a front-end that utilizes different sensor modalities to generate pose constraints (odometry and loop closures) and a back-end that incorporates these constraints to generate the optimized trajectory. This chapter describes the dimension-agnostic back-end (Sect. 2.1.1) developed by Dong et al. [5] and provides a straightforward extension to their front-end by transforming 3D scans into 2D scans (Sect. 2.2). Section 2.3 provides baseline results where the full 3D trajectory is estimated and a 2D map is built.

2.1 Problem Formulation

We introduce the multi-robot trajectory estimation as a Pose-SLAM problem and provide a brief overview. A more detailed discussion can be found in [5]. Pose-SLAM seeks to estimate the optimal history of poses of a robot given the constraints generated by sensor observations. Consecutive poses for a single robot are constrained by a combination of odometry (generated from wheel encoders or an IMU) and scan alignment between the consecutive scans (Iterative Closest Point (ICP)). Non-consecutive poses, including those of two different robots, can be constrained only through scan alignment. The joint probability distribution over the potential poses of the robots is parameterized as a factor-graph where the variables are poses and the factors are pose constraints. Therefore, for a robot r, the likelihood of a trajectory X^r given its observations Z^r is proportional to a prior on the initial pose and the product of each constraint $u_{i,j}^r$ between pose i and j is expressed as:

$$P(X^r \mid Z^r) \propto p(x_0^r) \prod_{i,j}^n p(u_{i,j}^r \mid x_i^r, x_j^r),$$
(2.1)

where $x_i^r \in X^r$ is the pose of the robot in its local frame at time t_i and $p(x_0^r)$ is a prior on the initial position. Given this factorization, the estimated trajectory \hat{X}^r is the input trajectory X^r that maximizes Equation 2.1:

$$\hat{X}^r = \operatorname*{argmax}_{X^r} p(X^r | Z^r).$$
(2.2)

The estimation of the trajectory likelihood is extended to multiple robots by including multi-robot data correspondences F as additional factors in the graph. Since each multi-robot factor involves only two robots, without loss of generality the remaining analysis considers only the interaction between two robots r_1 and r_2 . Each multi-robot correspondence is a transform represented as $u_{k,l}^{r_1,r_2} \in F$ for robot r_1 at time t_k and robot r_2 at time t_l and is computed from sensor observations $z_k^{r_1} \in Z^{r_1}$ and $z_l^{r_2} \in Z^{r_2}$. The transform induces a constraint on the poses of each robot. Using the factor-graph formulation, the likelihood of the full trajectory \mathcal{X} given all sensor measurements \mathcal{Z} is proportional to the likelihood of each of the individual factor graphs multiplied by constraints of the multi-robot factors:

$$P(\mathcal{X} \mid Z) \propto P(X^{r_1} \mid Z^{r_1}) \cdot P(X^{r_2} \mid Z^{r_2 2}) \cdot \prod_{\substack{u_{k,l}^{r_1, r_2} \in F}} P(u_{k,l}^{r_1, r_2} \mid x_k^{r_1}, x_l^{r_2}).$$
(2.3)

In practice, multi-robot correspondences are not always accurate and a false multirobot factor would cause large errors in trajectory estimates. To determine the true inlier correspondences, an EM approach is formulated in the next section.



Figure 2.1: A potential relative transform between world frame's $(\hat{T}_{r_1}^{r_2})$ is estimated by a scan correspondence (yellow) between robot r_1 at position $x_k^{r_1}$ (blue) and robot r_2 at position $x_l^{r_2}$ (purple).

2.1.1 Expectation Maximization Formulation for

Relative Transform Estimation

In order to determine which constraints in F are true correspondences, an Expectation Maximization (EM) is used as follows. Note that each multi-robot pose constraint $u_{k,l}^{r_1,r_2}$ leads to an estimate of the relative transform between $x_0^{r_1}$ and $x_0^{r_2}$. The relative transform can be computed by composing the scan registration pose constraint and the pose of robot r_1 and then expressing this composition in the frame of robot r_2 as seen in Fig. 2.1.

The estimated transform between r_1 and r_2 's world frame, $\hat{T}_{r_1}^{r_2}$, can be expressed as :

$$\hat{T}_{r_1}^{r_2} = \left(x_k^{r_1} \oplus u_{k,l}^{r_1,r_2} \right) \oplus x_l^{r_2}, \tag{2.4}$$

where \oplus is the compose operator and the notation $a \ominus b$ for any two poses a, b is used to express b in the local frame of a [17]. In future sections, since only two robots are considered, the superscript and subscript will be omitted for the relative transform $\hat{T}_{r_1}^{r_2}$.

The key assumption here is that inlier relative transform estimates will cluster whereas outliers will be distributed randomly as shown in Fig. 2.2. Expectation Maximization will be used to identify clustered measurements and determine the most likely transform. In order to characterize whether a scan is an inlier or outlier, a set of latent binary variables $j \in J$ are introduced to correspond with each multi-robot factor such that j = 1 if the related multi-robot factor is an inlier and 0 otherwise. In order to determine the binary variables J, the expectation maximization concurrently estimates a relative transform \hat{T} and determines the set of inlier scan matches. Given a set of inlier scan matches, a relative transform may be estimated by taking the average of the relative transforms estimated by each scan match. Further, when given a relative transform, one can select only those scan correspondences that provide a similar estimate. The most likely overall relative transform \hat{T} is therefore the scan correspondence that has the highest probability across all potential inlier sets given the current set of observations and the estimated trajectory:

$$\hat{T} = \arg\max_{T} \sum_{J} p(T, J \mid \hat{X}, Z).$$
(2.5)

EM is an iterative approach that outputs an estimated transform \hat{T} in every iteration. The two steps of each iteration of the EM optimization are written as:

1. **E step:** Calculate the lower bound on the probability of a transform T given estimated trajectory \hat{X} and observations Z by iterating over the set of potential classifications J:

$$Q(T \mid \hat{T}^{t}) = \sum_{J} p(J \mid \hat{T}^{t}, \hat{X}, Z) \log \left[p(T, J \mid \hat{X}, Z) \right].$$
(2.6)

2. M step: The transform that maximizes the lower bound on the probability of that transform given the previous transform estimate:

$$\hat{T}^{t+1} = \arg\max_{T} Q(T \mid \hat{T}^t).$$
(2.7)



Figure 2.2: Each line represents a 2D relative transform where the position of the arrow represents the delta translation and the heading represents the delta yaw. The relative transforms estimated by inlier scan matches are clustered (black circle) while outlier scan matches are randomly distributed.

The expectation step can be solved assuming that the binary variables indicating the inliers are statistically independent [5]. A fixed transform \hat{T}^t provides a distribution over J. The expectation over this distribution provides a likelihood value that is used to find \hat{T}^{t+1} . EM performs a clustering on the relative transform estimates with the assumption that true inlier scan matches lead to a consistent relative transform between a pair of robot's local frames. As the EM algorithm is sensitive to initialization, simpler clustering methods such as k-means are used to provide starting points. Once the relative transforms have been clustered by k-means, they are used as initial estimates in the EM and refined to obtain a final relative transform between each pair of robots.

2.2 System Details and Naïve 3D Extension

The EM formulation presented above makes no assumptions on how constraints were generated and is readily extended to 3D. We now discuss the extension of the front-end from the 2D approach [5] to a 3D approach that generates 3D scan correspondence constraints with panoramic sensors like the Velodyne-16 Puck. The goal of the frontend is to generate multi-robot loop closures by attempting to match received remote scans with local scans.

The naïve extension includes a pre-processing step that creates a horizontal 2D slice from a 3D scan. The slice is obtained by first pruning all points above and below a certain z value and simply setting the z value of all remaining points to 0. This aligned 2D scan can then simply be processed as described in [5]. We provide a brief overview of the loop closure detection process here.

The 2D scan is decomposed into features using Fast Laser Interest Region Transform's (FLIRT's) curvature detector and beta grid descriptors [31]. The extracted descriptors from the local scans are stored in a kd-tree for efficient nearest-neighbor search in the future. Since the data is high-dimensional, FLANN [21] is utilized to store the kd-tree.

When a remote scan is received, it is decomposed into FLIRT descriptors and

each descriptor is queried against the kd-tree to extract the closest K descriptors and their associated local scan indicies. The returned local scan indexes are placed into a histogram where each bin counts the number of features shared with the query scan. An example of a histogram can be seen in Fig. 2.3.



Figure 2.3: The histogram of matched features over scans identifies potential scan matches [5]. The query scan (blue) matches closely with the local scan (red) at the histogram peak index indicated by the pink arrow.

The histogram peaks are chosen as potential matches as they share a large number of features with the remote scan. A Random Sample Consensus (RANSAC) based feature association process is used to align the chosen local scan and the query scan and filter matches with insufficient consistent feature correspondences. ICP is then run to generate a refined transform $T_{l,k}^{2D}$ and fitness score *e* based on residual error. If the error from ICP is below a given threshold ($e < \lambda_{th}$), the calculated transform is used to generate a multi-robot constraint. The registration of 2D scans provides an x, y translation and yaw rotation. To obtain the full 3D transform from the scan registration, we assume the missing degrees of freedom (roll, pitch, and height) are initially consistent across all robots and use the roll, pitch, and height from the relative transform between their local poses. Let $T_k^{\prime r_1}$ and $T_l^{\prime r_2}$ be the transforms incorporating only the roll, pitch, and height components of the local poses of robot r_1 at time kand r_2 at time l respectively. If the 2D scan registration provides a transform $T_{l,k}^{2d}$, the full 3D relative pose constraint is:

$$u_{k,l}^{r_1,r_2} = \left(T_k^{\prime r_1}\right)^{-1} \cdot T_{l,k}^{2d} \cdot T_l^{\prime r_2}.$$
(2.8)

The roll, pitch, and height can be measured independently for each scan so no significant drift occurs in these components.

2.3 Results

We present here preliminary results generated by taking 2D slices of 3D scans and using FLIRT features and descriptors. In this experiment, three aerial robots map a virtual building with a simulated Velodyne-16 sensor. We inject Gaussian noise into the state provided by the simulator to provide a better replication of real-world data. The full map of the environment generated from the three robots can be seen in Fig. 2.4.

At initialization, the robots are unaware of the starting position of the other robots. While the relative transform calculation is correctly estimated, the level of data usage even for 2D scans surpass our given 150 KB/s limit. Furthermore, this environment has distinct features and minimal perceptual aliasing. The 2.5D environment is not amenable to 3D features, however many real world environments with


Figure 2.4: The estimated trajectory of the 3 robots (red, green, blue) with the joint map built by the green robot. This run assumes full communication amongst the agents.

clutter require the use of 3D data and features. In the next chapter we introduce a communication strategy that allows us to share 3D data at high resolution while operating over bandwidth constrained networks. Furthermore, we complete the extension to full 3D in Chapter 4.

Chapter 3

Communication Constrained Information Routing

Multi-robot systems potentially offer substantial performance gains as compared to single robot systems in cooperative tasks such as exploration. However, in order to realize increased performance, team members must co-ordinate their decision making processes. For effective coordination, robots must share local information and process the shared data to estimate a global state. In our target scenarios of operating in communication constrained environments, individual robots share local sensor observations and state information to allow each robot to estimate the global map. We employ an ad-hoc mesh network utilizing the BATMAN protocol [23] in order to operate in unexplored environments without existing communication infrastructure. This chapter addresses operating within the bandwidth limitations of ad-hoc communication networks that inhibit robots from sharing all local sensor observations.

Naïve information sharing formulations in applications with information-rich sensors, such as RGB-D sensors, quickly saturate available network capacity [12] which impacts coordination across team members [14]. While RGB-D sensors and similar 3D sensors produce depth data around 10-100MB/s, the observed network capacity of ad-hoc mesh networks under realistic situations is typically only around 0.1-1MB/s. Network capacity is additionally highly influenced by the distance between the robots and by the environment in which they operate [30]. In this chapter we present an approach that enables decentralized coordination through a distributed information sharing framework. Our approach adapts message passing to available bandwidth capacity in dynamically changing network topologies and prioritizes sharing information according to predicted utility at the recipient.

Relative transforms between robot reference frames are necessary for robots to be able to accumulate shared measurements into their local maps. We therefore develop a novel evaluation of scan utility by observing the expected impact of a scan's information on the calculation of relative transforms between robots using the distributed mapping formulation described in Chapter 2. Robots extract lightweight features of local scans and transmit these features to neighboring robots. A robot is able to compare any received features against the features of scans in its own scan database to rank local scans by the number of shared features since scans that share a significant number of features are likely to have overlapping observations. The generation of a relative transform between robots relies on identifying these corresponding observations between the robots, otherwise known as multi-robot loop closures. Our proposed scan utility function is the first to consider the likelihood of a scan to generate multi-robot loop closures which encourages sharing scans that enable relative transform calculations. Robots use an offer-request paradigm to communicate ranking metrics, allowing a pair of robots to determine the maximum number of mostinformative scans they are able to transmit between them, allowing them to share the most relevant measurements within network bandwidth limits.

While other approaches in the literature have used the information gain of a local

scan with respect to the corresponding robot's local map to estimate scan utility, our proposed approach, Communication Constrained Information Routing (CCIR), is the first approach to use virtual sensor readings to estimate the effect of a scan on another robot's local map prior to receiving the full scan. This "virtual" sensor is what allows robots to estimate the utility of measurements taken by neighboring robots and therefore negotiate optimal information flow. We proceed by discussing our methodology and provide extensive simulation results to compare our approach to fixed-rate down-selection. To validate the simulation results, we perform field experiments with two ground robots communicating over an ad-hoc network. While an approach of naïvely down-selecting sensor scans in order to remain within communication limits would miss key scans required to generate multi-robot loop closures, actively selecting scans based on their expected utility allows the robots to consistently generate relative transforms between each other and develop rich 3D maps while communicating across a severely constrained network.

3.1 Overview

To explain our approach for efficient distributed information gathering, we consider one side of the interaction between two robots i and j. We outline the six principal components governing measurements sharing from robot j to robot i below. The numbered steps directly correlate to the numbered arrows in the block diagram illustrated in Fig. 3.1.

- 1. First, robot i extracts features from its local observations and transmits these features to robot j.
- 2. Second, robot j compares the received features against its own local database

of scan features to estimate the likelihood of robot i's scan sharing observations with any scan in robot j's local database.

- 3. Third, robot j additionally evaluates its set of local scans to determine the information each scan contributes to robot j's map.
- 4. Fourth, robot j uses a combination of the loop closure potential (generated with information calculated in step 2) and information gain with respect to robot j's map (as determined in step 3) to rank its set of local scans by estimated utility to robot i . Using this ranked list of local scans, robot j shares with robot i information about at most N of robot j's highest ranking scans through an offer comprising each scan's location and number of shared features.
- 5. Fifth, robot i sorts the received scan offers from robot j using the expected impact of robot j's scan on robot i's own local map and potential for loop closure. Robot i then requests at most the N highest ranked scans from robot j, where N depends on the available capacity of the network connection between robot i and j.
- 6. Sixth, robot j sends scans requested by robot i to robot i.

In the following sections, we first explain the ranking function used to estimate scan utility using the loop closure potential and expected map information (Sect. 3.2.1). Sect. 3.2.2 further explains components 1 and 2 and Sect. 3.2.3 provides additional detail on components 3 and 5. Furthermore, Sect. 3.2.4 explains the offer-request framework used in components 4,5, and 6 to determine information routing.

3.2 Communication Constrained Information Routing

To most efficiently use available communication and computation resources, we match network utilization to the available network capacity. In variable network topologies, convergence to the optimal network routing is not feasible in real-time. To overcome this limitation, robots repeatedly adjust the number of offered, requested, and sent measurements under the offer-request paradigm as explained in Sect. 3.2.4. While this results in redundant offers and requests, the size of the offer and request messages are sufficiently small that the corresponding overhead is negligible. While this does not solve the global information routing problem, it minimizes the inter-robot negotiation required to establish information flow that bottlenecked Kassir et al's approach [14]. The inter-robot information flow is outlined in Figure 3.1.

3.2.1 Measurement Evaluation

In order for a robot j to determine a subset of measurements to offer to robot i, robot j considers the following two competing scenarios. First, robot j's observations that significantly overlap with robot i's observations have little value in terms of lowering the uncertainty of the map; however, these shared observations are extremely useful for determining loop closures and consequently for estimating the relative transform [5] between the communicating robots. Conversely, robot j's measurements that do not overlap with robot i's local measurements are useful for robot i to learn about the environment [13], but have little value for generating loop closures. Therefore we use a weighted cost function over these two measures to determine the communication value of a scan.



Figure 3.1: The diagram describes the one-sided relationship between a requesting robot i and provider robot j in the offer-response-request framework. Each robot is both a requester and provider and so will have all four sub-systems shown above. The gray box represents the distributed mapping component while the purple box represents the measurement accumulator. Sect. 3.1 contains a description of the information communicated in each numbered arrow.

The *utility* for a scan $S_i^{t'}$ from robot j at time t', to robot i at time t is defined as:

$$U(S_j^{t'}) = \lambda_m \cdot F(S_j^{t'}) + (1 - \lambda_m) \cdot I(S_j^{t'}), \qquad (3.1)$$

where $F(S_j^{t'})$ is the utility derived from the potential multi-robot loop closures as defined in Section 3.2.2 and $I(S_j^{t'})$ is the utility derived from the map uncertainty reduction achieved by the measurement as discussed in Section 3.2.3. We define λ_m as a parameter controlling the inherent trade off between generating loop closure and map information gain. A larger λ_m results in more overlapping scans and thus leads to lower uncertainty about the state whereas a smaller λ_m leads to more independent sensor readings and a lower uncertainty about the environment. We set λ_m to reliably generate relative transforms between robots and prioritize for unique sensor readings. This value will be dependent on the overlap of the robot trajectories, the environment they operate in, and the desired bound on state uncertainty.

The following sections discuss the individual components of Equation 3.1 in more detail.

3.2.2 Evaluating Loop Closure Potential

Loop closures between two robots are key components in generating a relative transform between the robots frames. Therefore, when two robots share scans that have a high likelihood of generating loop closures, the robots increase their probability of estimating their relative transform. This section describes how robots evaluate likelihood of detecting loop closures and use this likelihood as part of the ranking function.

When robot *i* locally generates a new measurement S_i^t at time *t*, it builds and sends a *feature message* to robot *j* (Arrow 1 in Fig. 3.1) containing the features $f_i^t = f(S_i^t)$ extracted from scan S_i^t . Each robot maintains a Kd-tree of locally generated measurements, where each measurement is stored according to its feature vector extracted using FLIRT's beta grid descriptor [31]. When a feature message is received at robot *j*, robot *j* queries its local Kd-tree to generate a list of scans with features similar to those in f_i^t (Arrow 2 in Fig. 3.1). Robot *j*'s scans which share similar features are ranked by the number of feature correspondences c_k and placed into a set:

 $\{S_j^{t_k} \mid \text{number of shared features between } f_j^{t_k} \text{and} f_i^t > c_{k,min}\},\$

ordered by increasing c_k where $c_{k,min}$ is a parameter describing the required minimum

correspondence count. The number of correspondences is a strong indicator of the similarity between two measurements and, therefore, the likelihood of a scan match.

In addition to improving the reliability of a relative transform calculation, an increased number of loop closures also reduces a robot's state uncertainty as each loop closure generates additional constraints in the pose graph. We use the product of the eigenvalues of robot j's marginal covariance matrix [4] at time t', $u_j^{t'}$, as a measure of robot j's state uncertainty. As the uncertainty of the pose estimate increases, robots weight scans which are useful for generating loop closures higher. Prior to generating a relative transform, the uncertainty value is set to some max value to encourage additional detection of loop closures.

Combining c_k (a scan's feature correspondence count) with $u_j^{t'}$ (a measure of robot j's pose uncertainty), the utility function is defined as:

$$F(S_j^{t_k}) = \frac{c_k u_j^{t'}}{\eta},$$
(3.2)

where η is a normalization constant. As an implementation note, since the utility function must be evaluated over all scans, the loop closure utility for measurements not selected from the local kd-tree is set to zero.

3.2.3 Evaluating Map Information Potential

Robots share measurements in order to reduce the uncertainty in each robot's map. However, the information content in a measurement relative to a given map depends on the set of measurements used to build that map, meaning a given measurement carries a different amount of information for each robot. We approximate the information content of robot j's local measurement with respect to robot i by considering both the measurement's local impact to robot j's map and the expected impact to robot i's map. The sending robot j evaluates a measurement by first considering the measurement relative to its own local measurement set in order to prune measurements from its database which have low information content or are redundant. When robot i receives an offer containing the position of robot j's local measurement, robot i considers the measurement's expected impact by using a simulated sensor reading at that location.

We first detail robot j's local evaluation of map information. Ideally, robot j would seek to identify the N measurement subset of its set of all measurements that would achieve the maximum combined entropy reduction for robot i. This is a variant of the Knapsack Problem with non-constant object costs, which has been shown to be NPhard. We therefore formulate an approximate approach by considering a sequential greedy ranking of the measurements. This has been shown to be within a factor of two of optimal [16]. At each step in the sequential algorithm, robot j selects and adds to its sorted list Z_j^i the measurement S_j^t that maximizes $MI(\hat{B}_j; S_j^t)$, the Mutual Information (MI) between the measurement and robot j's global environment belief \hat{B}_j . The belief $\hat{B}_j = f(Z_j^i \cup Y_j^i)$ is a function of the set of measurements comprising the sorted list Z_j^i and the set of measurements Y_j^i known by robot j to have been received by robot i. Robot j learns that robot i has received the measurement when robot i responds to robot j with confirmation of receipt.

Before a robot can integrate measurements received from other robots, a relative transform between the robots' reference frames must be established. To discourage sharing scans that do not contain shared features before a relative transform is built between robots i and j, a binary weighting factor RelTransformBuilt(i, j) which has value 1 if the relative pose is built and some small positive constant k > 0 otherwise is applied to the mutual information calculation. We select a non-zero constant to ensure robots utilize available capacity when there are no measurements with shared features but there are measurements with high information content. The modified information function follows as:

$$I(S_j^{i'}) = \mathrm{MI}(\hat{B}^j; S_j^{i'}) \cdot \mathrm{RelTransformBuilt}(i, j).$$
(3.3)

The measurement set Z_j^i is updated according to the rule:

$$Z_j^{i+} = Z_j^{i-} \cup \operatorname{argmax}_{S_j^{t'}} U(S_j^{t'}), \qquad (3.4)$$

Furthermore, robot *i* evaluates map information for an offered measurement $S_j^{t'}$ using the shared pose of robot *j* at time *t'*, p_j^t . Since robot *i* does not yet have scan $S_j^{t'}$, robot *i* approximates scan $S_j^{t'}$'s utility through a simulated measurement and application of Cauchy-Schwarz Quadratic Mutual Information (CSQMI) [2] between its local map and the simulated measurement. Note that robot *i* is only able to generate a simulated sensor measurement at robot *j*'s position at time *t'* if robot *i* already knows the relative transform between robots *i* and *j*. If the relative transform is unknown, the map information of $S_j^{t'}$ is set to a small constant k > 0 as before.

3.2.4 Offer-request Paradigm

Depending on available network capacity between robots i and j, robot j determines the number of scans (at most N) that are of potential value to robot i. Using the update rule for the measurement set (Eq. 3.4), robot j will continue sorting measurements until set Z_j^i has cardinality N, $|Z_j^i| = N$, or there are no more measurements to consider. Given the set of robots C_j that robot j can currently communicate with, robot j periodically constructs an offer with information about measurements Z_j^i for each $i \in C_j$; the maximum size of each measurement set Z_j^i is

$$N = \frac{\text{TotalBandwidth}}{(\text{AverageScanSize} \cdot |C_i|)}$$

Robot j sends robot i information about each scan in the sorted set Z_j^i including the scan's local position and the number of shared features between the scan's features and features received from robot i. The system diagram (Fig. 3.1) depicts the offer message as the arrow labeled four.

Robots respond to received offers at a fixed rate and therefore will often need to respond to offers from multiple robots at a single time. As a result, available network capacity may make it infeasible for robot *i* to request all offered measurements: $M = \bigcup_{j \in O_i} Z_k^i$, where O_i is the set of robots that have sent robot *i* an offer. Robot *i* uses the CSQMI rewards (described previously in Sect. 3.2.3) in conjunction with robot *i*'s own pose uncertainty to evaluate the offered scans and sequentially sort the set of offered measurements M into a set of requests R_i^j where $|R_i^j| \leq N$. The requested measurements correspond to message five in the system diagram (Fig. 3.1). Robot *j* collects requests from its neighboring robots C_j and responds with the requested scans (message six in Fig. 3.1).

3.3 Results

The described framework reduces the network burden while preserving the quality of the distributed map as compared to approaches that down-select messages at fixed intervals. The construction of relative transforms between robots impacts the ability of robots to incorporate foreign measurements into their local map representation, significantly impacting the rate of entropy reduction across the distributed map. We therefore quantify the performance of our approach in terms of the uncertainty reduced in the distributed map on each robot and the calculation of the relative transforms between robots.

3.3.1 Implementation Details and Complexity Analysis

This section describes several implementation details and analyzes the run-time complexity of the proposed approach. Robots utilize an occupancy grid map to represent their environment. When a robot receives a new scan, the robot performs ray-casting to determine which 3D cells in the occupancy grid are affected by the sensor reading. In each of these cells, a constant time operation is performed to calculate the change in uncertainty of that cell based on a sensor model. Let P be the points in the scan, R, the cell size, and C, the number of cells affected by the scan. Note that C is inversely proportional to R since larger cells imply a fewer number of cells for a given area. On compute constrained systems, the value of R can be increased to decrease the number of cells and therefore minimize the required computation. Since the sorting process iterates over the full list of size L of a robot's local scans until it has selected N scans, the sorting process calculates scan value $\mathcal{O}(NL)$ times. In order to rank its set of scans that have not been shared yet based on map information utility, a robot initializes a temporary map containing scans that the robot estimates are known to all other robots on the team. This allows the sorting algorithm to evaluate the Nbest scans to transmit with respect to the local robot's estimate of the team's total knowledge. To evaluate a scan, the sorting algorithm calculates the change in entropy using the sensor model for each cell affected by the sensor. After each iteration over the full list of scans, the sorting algorithm adds the measurement with the largest change in entropy to the temporary map. Sorting is on the order of $\mathcal{O}\left(NL\hat{C}\right)$ where \hat{C} is the average number of cells in a scan. An initial thresholding based on scan information can be used to remove scans from the sorting list. In practice, a resolution of 0.1 meters achieved real-time performance on the ground robot seen in Fig. 3.8.

3.3.2 Comparison with Fixed-Rate Down-selection

We evaluate our approach in a large un-cluttered 2.5D building and cave as well as a cluttered room environment. In all experiments, robots follow preset trajectories so that the competing approaches can be directly compared. Using hardware experiments, we set the network bandwidth to 150 KB/s to ensure under-saturation of the network.



Figure 3.2: The different environments used for simulation, including the 2.5D extrusions of (a) a building and (b) a cave where Velodyne-16 scans are simulated as well as a 3D environment with (c) clutter.

While it is possible to down-sample scans at a rate that limits saturation, the performance is poor both in the ability to generate relative transforms and map completeness as seen in Figure 3.3. To characterize the operation of CCIR, we select the minimal fixed-rate down-selection rate that results in the successful estimation of relative transforms between robots and a distributed map with a final volume within 10% of the distributed map constructed with CCIR. This allows us to describe the network savings when using CCIR as compared to a similarly performing fixed-rate



Figure 3.3: (a) 3D map generated by using CCIR, (b) the map volumes over time, and (c) the network utilizations of three robots using CCIR and fixed-rate down-selection. Selecting a fixed-rate down-sample amenable to the bandwidth supported significantly degrades the completeness of the generated map. The final volume for the naïve approach is 36% of the volume of the map generated using CCIR.

down-selection approach.

3.3.3 Simulation Results

The simulation experiments include a variable number of robots operating with simulated 3D depth cameras and 2D Lidar scanners communicating over an artificial network. The artificial network logged the data transmitted but did not impose any latency or bandwidth limitations. As our approach is agnostic to the type of depth sensors, we additionally provide tests with simulated Velodyne-16 scans. A subset of 3D Velodyne scans are compressed to 2D as explained in Sect. 2.3 to allow FLIRT features to be used for scan correspondence detection. Furthermore, we validate the simulation results with two ground robots, pictured in Fig. 3.8. In each environment tested, our proposed approach, CCIR, outperforms the fixed-rate down-selection approach by sharing less data while reliably estimating relative transforms and reducing uncertainty in the map at an equivalent rate. Figure 3.4 shows the comparison of CCIR and the fixed rate approach using the building environment. The entropy reduction over time remains similar through the run while the required data transmission is 169% higher for fixed-rate approaches. Two tests with three robots each were conducted in the cluttered environment. As can be seen from Fig. 3.5, CCIR respects the set bandwidth while constructing a map similar to a fixed rate approach that utilized 292% and 381% of the bandwidth limit respectively for each experiment. The clutter causes more irregularity in scan utility as compared to simpler environments and so the bandwidth required by a fixed-rate approach is higher than in the prior experiment.

We also observed the effect of varying the number of robots in the system. While



Figure 3.4: The entropy reduced over time (a) for each approach is comparable while the cumulative network utilization (b) for CCIR is 59% of the that of fixed-rate downselection. The dark line represents the average while the shaded region represents three standard deviations.



Figure 3.5: Results from two different runs of three different trajectories through the more complicated environment. The figure contains two distributed maps (a), (d) as well as the mean and variance in reduction of entropy over time (b), (e) and the network utilization (c), (f). The colored arrows represent locations where scans are taken. Red represents local 2D scans, yellow is local 3D scans, dark blue is remote 2D scans, and cyan arrows are foreign 3D scans.

our approach is useful for smaller teams of around 3 robots, we see that it provides even more improvement as team size increases, as illustrated using a team of 6 robots



Figure 3.6: Comparison of CCIR and fixed-rate down-selection with six robots in cluttered environment. The figure includes (a) the entropy reduction over time and (b) the network usage as well as (c),(d) the generated maps.

as shown in Fig. 3.6.

We show the scalability of our approach over 3, 6, and 9 robots operating in the 2.5D cave environment. The 150 KB/s limit required by the mesh network was maintained until 9 robots operated concurrently. Due to the configuration and overlap in their trajectories, the team of 9 robots initially required a higher bandwidth in order to generate relative transforms between their world frames. However, this assumed every pair of robots maintained communication throughout the experiment. If robots



Figure 3.7: (a) The entropy reduction over time for a team of 9 robots operating with a Velodyne is comparatively the same between the two approaches while the network utilization (b) for CCIR is 39.5% of the network utilization of fixed-rate downselection. The required bandwidth was increased to 200 KB/s to allow convergence of the relative transform calculation.

Table 3.1: Comparison of CCIR and fixed-rate down-selection over various numbers of robots.

	CCIR	Fixed-Rate
3 Robots	150 KB/s	260 KB/s
6 Robots	150 KB/s	440 KB/s
9 Robots	200 KB/s	495 KB/s

communicate with a subset of the robots, then the network limitation of 150 KB/s still holds. Figure 3.7 presents results from 9 robots tested using a simulated Velodyne sensor. Table 3.1 shows the required bandwidths over the various runs.

3.3.4 Field Experiments

We additionally performed experiments on hardware platforms to validate the results seen in simulation. The ground robot is pictured in Fig 3.8. It is equipped with a Gigabyte Brix computer a mobile i7 processor and 16 GB RAM, a Bullet 1 Watt network interface, a Hokyu U30MLX lidar scanner and Orbecc Astra camera. One of the major shortcomings of Kassir et al.'s [14] approach is the lack of real-time testing



Figure 3.8: The robot used for hardware tests is equipped with a bullet wireless transmitter, a Hokuyo 2D Lidar, an Orbecc Astra RGB-D camera and a core i7 Gigabyte Brix for compute.

on systems with large latency, which is a significant factor affecting the performance of this approach. For our hardware experiments, the latency of the system was between 50-100ms for a single message between two robots; however, robots need only an offer and response to share a scan rather than a distributed optimization, thus, the latency did not severely impact mapping performance of the team of robots.

In Figure 3.9, two robots have constructed a 3D map while communicating over a mesh network. To calculate the required network usage for a fixed-rate down-selection approach, the collected depth and Lidar data was post-processed using the simulation setup described in Sect. 3.3.3. Even with the latency observed between systems, CCIR is able to perform similarly to simulation on real-hardware, thereby validating the prior simulation experiments. By remaining below the prescribed bandwidth limit,



Figure 3.9: (a) Map constructed from two ground robots using 2D Lidar and depth camera. (b) Network usage stays below set bandwidth requirements while fixed-rate down-selection requires additional bandwidth to construct an equivalent map.

the robots have minimized network usage and thereby allowed more bandwidth for coordination and shortened delays between shared information.

3.4 Chapter Summary

In this chapter, we provided an overview of a distributed information gathering strategy that was amenable to severely constrained communication networks. When dealing with high-data sensors such as depth cameras or 3D Lidars, not all sensor scans can be shared. Fixed-rate down-selection leads to a significant loss of information in these scenarios, and we show an intelligent strategy for determining the routing of scans based on a novel utility measurement which dramatically out performs the fixed-rate down-selection baseline. We develop this utility measurement by observing the impact of a scan on the global state and distributed map and use the resulting utility measure to selectively offer and request scans that have high predicted impact. Our results show that our approach of CCIR enables a team of mobile robots sharing scans over a bandwidth-limited wireless mesh network to construct detailed and accurate maps of the environment.

Chapter 4

Extending the Distributed Mapping Framework to 3D

This chapter outlines the extension of the distributed mapping process to 3D sensors and the methods to deal with the resulting perceptual aliasing in repetitive environments that occurs due to inherent problems with 3D feature detection. In this chapter, we make use of a quadrotor equipped with a Velodyne-16 LIDAR scanner. While the Velodyne sensor produces 300,000 points per second, points further than a few meters from the sensor are extremely sparse and require special attention. Iterative Closest Point (ICP) algorithms that utilize the point-to-point error minimizer do not produce accurate scan registrations with these scans since exact point to point correspondences are unlikely [26]. Sect. 4.1 provides an overview of the 3D features we employ for scan recognition and the modifications we make to allow for them to work on a sparse Velodyne-16 sensor in order to allow us to efficiently and robustly match scans. Sections 4.2 and 4.3 detail robustification techniques to help deal with higher levels of outlier scan correspondences. Using these methods, we attempt to run CCIR in our target mine environment and show that the algorithm fails due to high levels of perceptual aliasing. We conclude with post-processed results from hardware experiments involving two quadrotors operating in an underground mine.

4.1 3D Features

Features and their corresponding descriptors provide a succinct overview of the structure of a scan and are used for determining potential scan matches. The goal of feature selection is then to collect unique parts of a scan and provide reproduce-able descriptors from scans taken from locations close to the current scan but which are distinct from the descriptors of scans taken elsewhere. While work by Serafin et al. [27] looks at detecting features from sparse point clouds using planes and lines, environments such as mines (Fig. 1.2) are not amenable for such simplifications. One of the current state-of-the-art approaches for describing range measurements is Normal Aligned Radial Features (NARF) [29]. Acting directly on the range image, NARF finds areas that are both stable and have depth variance. Due to the lack of clutter in the target mine environment, depth disparities are restricted to pillar corners. As pillar corners are extremely repetitive, NARF performs poorly.

We have found that ISS Keypoints [32] in conjunction with Spin Images [11] as descriptors provide a reasonable characterization of scans in our repetitive, underground mine environment. These approaches, however, were not developed for sparse point clouds because keypoint and descriptor calculation rely on point normal calculations. Normals are vectors that describe the surface captured by a point cloud by providing the direction of the surface tangent (Fig. 4.1). In addition, normal vectors are estimated for each point by fitting a 3D Gaussian over the point and it's neighboring points within a fixed radius and extracting the eigenvector corresponding to the smallest eigenvalue. Intuitively this is the direction that is the most "flat" and this is therefore assumed to be the surface vector at that point.



Figure 4.1: A normal for a point is the tangent vector of the surface at the point. The blue circle denotes the support of the normal, i.e. the points it uses to estimate the Gaussian which it uses for point-normal calculation.

Normal vectors are often used because they are view-invariant and so descriptors based on normals will also be view-invariant. However, normals rely heavily on the density of points and the radius which defines their neighbors. Velodyne scans exaggerate this issue since scan density varies with respect to a point's distance from the sensor. In Fig 4.1 we illustrate how a given radius leads to degenerate normal vectors.

We handle sparse Velodyne point clouds in the non-cluttered environments of man made tunnels and mines by taking advantage of the enclosed nature of these environments. Excluding points on the ground and ceiling, the limited vertical range of the Velodyne coupled with the uncluttered tunnel environment allows us to make the assumption that accurate normal estimates will have z components close to 0 as seen in Fig. 4.2. By filtering out all normals and their corresponding points that have |z| > 0.5, we can show improved point correspondence detection (matching of



Figure 4.2: The unfiltered pointcloud side view (a) and top view (b). After removing normals that have a normalized z-value not within ± 0.5 , we obtain a set of sensible normals, shown in side view (c) and top view (d)

features from a scan with features of another scan) as depicted in Fig 4.3.

Feature detection is also dependent on scan resolution. CCIR minimizes the need for scan data reduction through down-sampling by sharing fewer scans. However, the size of a Velodyne scan (350Kb) is still significantly larger than the standard message packets (64Kb) used for communication. We wish to avoid splitting a single scan between too many different packets as the lossy nature of the communication could lead to frequently dropped packets and scan corruption. Therefore, we downsample the scan using a voxel grid filter with a cell size of 0.1m to reduce the scan size while not significantly degrading the calculation of features (Fig. 4.4). Robots now transmit the reduced scan (150KB) with three standard packets rather than six packets, thereby increasing data transmission reliability. Loss of unique features and descriptors due to the downsampling and repetitiveness of certain features increase incidents of perceptual aliasing as scans become less distinctive. The histogram approach used in Chapter 2 to detect correspondences results in many false detections



Figure 4.3: We show the feature correspondences between two scans with (a) and without (b) normals that are degenerated due to sparsity of the points. Yellow lines connect points with similar descriptors. While the scan size is reduced significantly, there are no more outlier matches.



Figure 4.4: Subfigures depict recognized features (green) in the same environment with successive downsampling: the full scan (a) along with voxel grid down-sampled scans using voxel sizes of 0.1 (b) and 0.2 (c). Downsampling reduces the ability of feature detectors to reliably detect interest points if the filter size is large.

in these situations as seen in Fig. 4.5. Therefore, we introduce robust methods to



Figure 4.5: Histogram of shared features over a robot's scan data base for a query scan (blue). The black bar represents the number of shared features between the query scan (blue) and the scan at the index given by x. The green bars represent scans that are within 3m and are considered ground truth matches. Note how similar false matches look.

mitigate aliasing effects in the following two sections (Sects. 4.2 and 4.3).

4.2 Fast-Appearance Based Matching

To improve the accuracy of correspondence detection, we utilize FAB-Map [23] instead of the histogram approach described in Chapter 2. The key idea of FAB-Map is to learn the co-dependence of scan features so as to make a more informed match detection. Using a Chow-Liu Tree, the approach attempts to approximate the full joint distribution over the probability of feature occurrence using pairwise conditional probabilities. Rather than assuming each feature occurs evenly and independently, FAB-Map learns the distribution through a training phase. More details on the approach can be found in [23].

Table 4.1 compares the Histogram and FAB-Map approaches for both 2D and 3D features in the mine environment. 3D features allow robots to generate a greater number of true positive estimates since they enable robots operating at different heights to more reliably find correspondences across heights. The FAB-Map approach does consistently better over the Histogram approach, reducing the outlier ratio from 99.1% to 98.1%.

Table 4.1: Comparison of FAB Matching and Histogram approach over three mine data sets.

	True Positives	False Positives
2D FLIRT Histogram	2	980
2D FLIRT FAB-Map	3	621
3D Spin Histogram	8	876
3D Spin FAB-Map	10	534

However, the number of outliers compared to inliers is significant and requires additional improvements to the optimization back-end to allow operation over longer durations. Therefore, we introduce switchable constraints and describe the implementation in Sect 4.3.

4.3 Switchable Factors

We add detected local loop closures and multi-robot loop closures after the determination of the relative transforms between robot reference frames as switchable constraints as in [1]. The approach described in [1] modifies the between-factor constraint to include a switch variable that acts as a weight on the factor between the poses it connects. The value of the switch variable is continuous between 0 and 1 and determines the likelihood that the constraint represents an inlier match. When the value is close to 1, the constraint is used in the optimization; when the value is close to 0, the constraint is effectively ignored in the optimization. A visualization of the factor and switch variable can be seen in Fig 4.6, reproduced from [1].



Figure 4.6: The pose graph pictured, reproduced from [1], contains switch variable $s_{2,i}$ and switch factor (yellow) that can robustly determine inliers and outliers.

The use of switch variables allows the back-end optimization to determine the true set of loop closures. While Dong et al. [5] present an incremental EM algorithm that decides if each new detected loop is an outlier or not, the algorithm is computationally expensive and prone to false detections over time. Conversely, using switchable constraints as described in [1] is fast, allows the system to modify what it believes to be the inlier set over time, naturally taking into account the uncertainty of the trajectory.

In Figure 4.7, we see the effect of adapting the switch variables prior factor and note how this affects the final pose graph optimization. Without a prior factor forcing the switch variable to a value of 1, the switch value would simply default to 0. Thus the strength of the prior factor determines how much to trust the loop closures given to the optimization. More details of this approach can be found in [1].



Figure 4.7: A low prior strength (a) causes all loop closures to be ignored while a high strength prior (c) causes the false positives to be considered as well. A prior tuned for this environment provides an optimized map (b).

4.4 **Results:** Mine Environment

Using the modifications described in Sects. 4.2 and 4.3, we evaluate the distributed mapping formulation using CCIR in highly repetitive environments. To characterize the effect of the intelligent sharing without having to strongly consider the overlap in trajectory information, we developed an experiment involving running one collected data set using two virtual robots that output the saved data at different times. In this experiment, one robot's scan and odometry data is collected for an eight minute run. The run is split into two overlapping sections and run independently so as to appear to be two independent robots. The overlap in trajectories enabled our approach to estimate relative transforms and the results can be seen in Fig. 4.8.

Evaluating a separate experiment where two different robots initially shared 30 seconds of their trajectory, we are again able to show not only the detection of the relative transform between the robot reference frames but also significant network savings due to CCIR. The required level of down-selection to reliably determine a relative transform is significantly higher due to the fact that perceptual aliasing creates a number of false matches. In both experiments, the estimated relative transforms



Figure 4.8: The generated map (a), the entropy levels over time (b), and the resulting network usage (c) from a two-robot trial in the mine environment.

for the fixed-rate down-selection and CCIR approaches are within 0.1m of the true relative transform.

4.5 Chapter Summary

Establishing robust operation in 3D environments requires timely and efficient processing of 3D sensor data. Features and descriptors provide sparse representations of 3D sensor information and enable loop closure detections through modification of FAB-Map. Currently established 3D feature keypoint detectors and descriptors work



Figure 4.9: The map generated from the fixed-rate down-selection approach (a) and CCIR (b). The two maps are nearly identical; CCIR is able to generate a relative transform using 13.6% of the bandwidth used by fixed-rate down-selection (c).

well for dense data but must be adapted by filtering out degenerate normals to be used reliably for correspondence detection in sparse 3D pointclouds such as those generated by the Velodyne. In addition to affecting keypoint detection, the sparsity of the points affects scan registration. Traditional ICP methods that minimize point-to-point errors do poorly when dealing with sparse data. Improving the reliability of feature detection and scan registration improves robot performance significantly, however the requirement of sharing information over the network, even when using CCIR, limits the system performance. We introduce FAB-Map and Switchable Constraints to generate reliable multi-robot loop closures and improve system performance, decreasing the outlier match rate from 99.8% to 98.1%. While the robust distributed mapping formulation has been shown to work given a high outlier rate, the structurally repetitive nature of the underground mine environment breaks the fundamental assumption that outliers are randomly distributed rather than clustered. In the experiments provided, our approach is able to reliably detect the relative transform between frames and generate globally consistent maps. However, there are cases where the outlier rate exceeds the capability of our approach; towards operation in such scenarios, the Appendix section of this thesis provides promising future work that enables efficient and accurate scan correspondence detection. Departing from the feature-based matching paradigm, we purse a dense strategy that both decreases outlier matches and enables compressed representation amenable to constrained communication.

Chapter 5

Conclusion

5.1 Summary

Efficient multi-robot operation requires that the robots are able to co-localize and develop a shared environmental model. With these capabilities, multi-robot systems can effectively determine trajectories or actions that best benefit the team. In the context of exploration, robots can determine optimal actions which result in the desired degree of sensor redundancy to allow for faster exploration and minimized measurement uncertainty. A shared environmental model that accurately accounts for multi-robot loop closures can provide a more consistent estimate of the environment than a single robot map. In this thesis we focus on robust distributed 3D mapping within difficult environments that exhibit minimal texture and repetitive structure and have no existing communication infrastructure, relying on robots to communicate through constrained ad-hoc mesh networks. The development of the distributed map across robots depends on the sharing of relevant information and the formation of inter-robot relative transform estimates. Where previous work has mainly utilized 2D sensors to develop distributed maps, we extend Indelman and Dong's distributed mapping framework [5] to work effectively with 3D sensors. Extending from 2D to 3D is complicated by the high data rate of the sensors and sparse nature of many 3D sensor measurements. These issues are further exacerbated when working in our low-feature, repetitive environments subject to perceptual aliasing. Being able to incorporate the additional information offered by 3D measurements, however, results in more detailed and accurate maps, especially for systems which maneuver in three dimensions such as quad-rotors. Furthermore, accurate relative transform estimates allow shared information to be fused correctly for efficient joint trajectory generation. With this novel distributed mapping framework, a team of robots is now able to effectively explore using a communication constrained network despite utilizing high data rate sensors.

In the second chapter we provide an introduction of the distributed mapping formulation developed by Dong et al. [5] and propose a naive extension to allow for 3D operation. Our work focuses on developing initial relative transforms via sharing local sensor information. While finding relative transforms is trivial when all areas of the environment are unique, perceptual aliasing causes incorrect data associations and complicates the estimation of relative transforms. Using an expectation maximization approach that assumes inlier scan correspondences lead to consistent relative transform estimates, robots are able to reliably generate transforms between their own and teammates' reference frames.

To mitigate the effects of network bandwidth constraints, we develop a scan utility and sharing framework that enables robots to send only the most useful scans to not exceed bandwidth limitations. Our utility function balances generating loop closures with reducing uncertainty in the map, enabling robots to reliably generate relative transforms between themselves as well as expand the area observed by each robot. In most environments, a fixed-rate down-selection strategy that attempts
to maintain bandwidth is unable to generate reliable transforms. Our approach, Communication Constrained Information Routing (CCIR), not only reliably generates relative transforms but also develops distributed maps of equal volume to fixed-rate strategies requiring 280% of the bandwidth limit. By enabling efficient operation that works within network saturation limits, CCIR ensures that coordination messages are transmitted in a timely manner and significantly reduces message dropouts due to network over-saturation.

Extending CCIR to utilize full 3D information in our target mine environment requires additional features and robust strategies to handle the increased levels of perceptual aliasing. Moving from FLIRT features to ISS Key-points and Spin Image histograms allows us to use 3D scans for observation correspondences. However, the repetitive nature of the mine environment is not amenable to the feature correspondence approach discussed in Chapter 2; we therefore incorporate the additional robust methodologies of FAB-Map [23] and Switchable Constraints [1] to reduce outlier ratios of 99.8% to 98.1% and enable reliable operation. Furthermore, given the rapid changes in texture observed in 3D scans given small motions relative to the environment, robots greatly benefit from intelligent down-selection to share relevant scans for loop closure detection. In the challenging mine environment, CCIR is able to generate a relative transform between two robots with 13.6% of the data transmission required by fixed-rate down-selection.

Despite the described advances of the CCIR algorithm, it is possible for the outlier ratio to exceed 99% under certain scenarios, showing that feature-based methods reach a fundamental limit. Another major complication of our required tunnel environment is that outlier correspondences are no longer scattered but appear grouped, thereby breaking one of the fundamental assumptions of the robust EM approach for calculating relative transforms. In order to effectively map in this environment, we require a model that captures the surface of the tunnel, is memory efficient enough to transmit, and accurate enough that we can uniquely identify different areas of the mine. To address this problem, we propose using a Gaussian Mixture Model that learns the distribution of the environment with high fidelity and minimal memory footprint [28]. Our initial results using this approach are presented in the Appendix section. While we are able to learn a Gaussian Mixture Model over the 3D point cloud, our current implementation of learning the GMM is not real-time viable and the computational cost of ICP scales poorly to longer duration experiments or to larger team sizes. However, our initial results show significant promise for future work.

5.2 Future Work

As previously mentioned, we look to incorporate the HGMM mapping framework into our distributed mapping formulation. A reliable detector of potential loop closures based on features extracted from the HGMM itself would help eliminate much of the work of ICP. Furthermore, developing and utilizing a registration process that works directly on the hierarchy of mixtures themselves rather than over 3D points has potential to be much faster as it operates over less data.

In addition to exploring the potential of the HGMM approach, another vein of research addresses scaling to larger teams. The scalability of the current implementation grows with the number of pair-wise connections. In order to allow concurrent operation of hundreds of robots, new strategies akin to artificial clustering and creation of hierarchies must be developed so that each robot at most handles communication with some fixed upper-bound of robots.

Finally, we are interested in pursuing a more reliable approach for scan matching

with sparse and non-uniform data. While some work has dealt with the sparsity of 3D points, a more general framework that does not rely on regular geometric environmental features such as lines and planes would open our approach to operation over a wide range of additional environments.

Chapter 6

Appendix

6.1 HGMM Models

A Gaussian Mixture model (GMM) is a powerful tool for modeling multi-modal distributions characteristic of the operating environment of an autonomous system. Given a point cloud Z of size N with points $z_i \in Z$, a robot generates a Gaussian Mixture model with J components where each component is specified by parameters $\theta_j = (\pi_j, \mu_j, \Sigma_j)$, and π_j, μ_j , and Σ_j are the prior, mean, and covariance matrix respectively for the *j*th component. The likelihood of an observed pointcloud to be generated by this Gaussian Mixture Model is given as:

$$p(Z \mid \theta) = \prod_{i=1}^{N} p(z_i \mid \theta)$$
(6.1)

$$= \prod_{i=1}^{N} \sum_{j=1}^{J} \pi_{j} p(z_{i} \mid \theta_{j}), \qquad (6.2)$$

where the probability of an observation z_i given the HGMM parameters θ_j is given by the normal distribution \mathcal{N} :

$$p(z_i \mid \theta_j) = \mathcal{N}(z_i \mid \theta_j), \tag{6.3}$$

and the corresponding log-likelihood of the data is:

$$\ln p(Z \mid \theta) = \sum_{i=1}^{N} \ln \sum_{j=1}^{J} \pi_{j} p(z_{i} \mid \theta_{j}).$$
 (6.4)

Srivastava et al [28] propose a methodology to learn a Hierarchy of Gaussian Mixtures (HGMM) as an approximation of the underlying distribution over occupied space in the environment. They show that this representation performs favorably in terms of memory-footprint and accuracy compared to other representations proposed in the literature ([8], [7], [24]). The proposed methodology estimates the number of HGMM components required for an accurate environment representation via an iterative procedure based on an information-theoretic measure. This enables the HGMM to provide a compressed, high-fidelity representation of the input sensor data. We refer the reader to [28] for a detailed discussion of the methodology.

6.2 Novel Loop Closure Detection

We use an HGMM representation to enable loop-closure detection in the challenging mine environment. Approaches that involve quantization of perceptual information including feature-based techniques such as Communication Constrained Information Routing (Chapter 3), suffer in environments that exhibit repetitive perceptual aliasing or lack of rich perceptual information. An HGMM representation avoids scan data quantization which enables the model to capture all the perceptual information made available by the sensors. This makes place-matching based on the HGMM representation more robust compared to approaches that quantize measurement information.

The methodology to detect a place-match involves an incremental generation of a map based on Hierarchical Gaussian Mixture models as outlined by [28] for incoming sensor data. A novelty criterion based on the likelihood of the incoming data to be represented by the existing model is used to define key-frames \mathcal{K} in the map. The generative capability of the GMM is leveraged when querying the map for a placematch. Specifically, the current point-cloud Z is compared to all the key-frames in \mathcal{K} to check for a potential place-match. For every key-frame, the environment region represented by that key-frame is regenerated by sampling from the corresponding GMM. Generalized Iterative Closest Point (GICP) [26] is then applied between the sampled point-cloud \mathcal{C} and Z to determine a potential place-match. The ability to store a high fidelity representation of the previously visited locations was not possible using CCIR due to the increasingly high memory requirements. However, the minimal memory footprint of the HGMM approach not only allows long term high quality storage but also allows efficient transmission.

The results of the loop closure detection over several experiments are presented in Table 6.1. The outlier rate of the HGMM-based loop closure detection method is 34% of that of feature-based methods. An example of a single robot operating on part of Mine 1 can be visualized in Fig. 6.1.

	TP (FAB)	FP (FAB)	TP (HGMM)	FP (HGMM)
Indoor Building	4	20	16	0
Mine 1	6	78	71	19
Mine 2	2	140	4	1
Mine 3	2	316	5	252

Table 6.1: Comparison of HGMM-based approach and FAB Matching.



Figure 6.1: The full map (a) contrasted with the pointcloud reconstructed by sampling from final HGMM (b). The trajectory of the robot is shown in black and the generated loop closures are shown in yellow. The HGMM-based approach generated fewer false matches (d) than the feature-based approach (c).

The current process of utilizing HGMM is currently not real-time viable due to the complexity in generating the HGMM and, more importantly, the large number of GICP registrations. Towards enabling faster operation, we are currently pursuing a parallelized approach to generating HGMMs as well as a robust pointcloud and GMM registration method as proposed by Eckart and Kelly [6].

Bibliography

- Pratik Agarwal, Gian Diego Tipaldi, Luciano Spinello, Cyrill Stachniss, and Wolfram Burgard. Robust map optimization using dynamic covariance scaling. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 62–69. IEEE, 2013.
- [2] Benjamin Charrow, Sikang Liu, Vijay Kumar, and Nathan Michael. Informationtheoretic mapping using cauchy-schwarz quadratic mutual information. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 4791–4798. IEEE, 2015.
- [3] Siddharth Choudhary, Luca Carlone, Carlos Nieto, John Rogers, Henrik I Christensen, and Frank Dellaert. Distributed mapping with privacy and communication constraints: Lightweight algorithms and object-based models. arXiv preprint arXiv:1702.03435, 2017.
- [4] Christophe Croux and Gentiane Haesbroeck. Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika*, pages 603–618, 2000.
- [5] Jing Dong, Erik Nelson, Vadim Indelman, Nathan Michael, and Frank Dellaert. Distributed real-time cooperative localization and mapping using an uncertainty-

aware expectation maximization approach. In *Robotics and Automation (ICRA)*, 2015 IEEE International Conference on, pages 5807–5814. IEEE, 2015.

- [6] Benjamin Eckart and Alonzo Kelly. Rem-seg: A robust em algorithm for parallel segmentation and registration of point clouds. In *Intelligent Robots and Systems* (IROS), 2013 IEEE/RSJ International Conference on, pages 4355–4362. IEEE, 2013.
- [7] Erik Einhorn and Horst-Michael Gross. Generic ndt mapping in dynamic environments and its application for lifelong slam. *Robotics and Autonomous Systems*, 69:28–39, 2015.
- [8] Wolfgang Hess, Damon Kohler, Holger Rapp, and Daniel Andor. Real-time loop closure in 2d lidar slam. In 2016 IEEE International Conference on Robotics and Automation (ICRA), pages 1271–1278, 2016.
- [9] G. A. Hollinger, S. Yerramalli, S. Singh, U. Mitra, and G. S. Sukhatme. Distributed data fusion for multirobot search. 31(1):55–66. ISSN 1552-3098. doi: 10.1109/TRO.2014.2378411.
- [10] Andrew Howard. Multi-robot simultaneous localization and mapping using particle filters. The International Journal of Robotics Research, 25(12):1243–1256, 2006.
- [11] Andrew E. Johnson and Martial Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on pattern analysis and machine intelligence*, 21(5):433–449, 1999.
- [12] Jangeun Jun and Mihail L Sichitiu. The nominal capacity of wireless mesh networks. *IEEE wireless communications*, 10(5):8–14, 2003.

- [13] A. Kassir, R. Fitch, and S. Sukkarieh. Communication-efficient motion coordination and data fusion in information gathering teams. In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 5258–5265, doi: 10.1109/IROS.2016.7759773.
- [14] Abdallah Kassir, Robert Fitch, and Salah Sukkarieh. Communication-aware information gathering with dynamic information flow. 34(2):173-200, . ISSN 0278-3649. doi: 10.1177/0278364914556911. URL http://dx.doi.org/10.1177/0278364914556911.
- [15] Been Kim, Michael Kaess, Luke Fletcher, John Leonard, Abraham Bachrach, Nicholas Roy, and Seth Teller. Multiple relative pose graphs for robust cooperative mapping. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 3185–3192. IEEE, 2010.
- [16] Henrik Kretzschmar, Cyrill Stachniss, and Giorgio Grisetti. Efficient information-theoretic graph pruning for graph-based slam with laser range finders. In 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 865–871. IEEE, 2011.
- [17] Feng Lu et al. Robot pose estimation in unknown environments by matching 2d range scans. In Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on, pages 935–938. IEEE, 1994.
- [18] Raj Madhavan, Kingsley Fregene, and Lynne E Parker. Distributed cooperative outdoor multirobot localization and mapping. *Autonomous Robots*, 17(1):23–39, 2004.

- [19] A. Makarenko, A. Brooks, S. Williams, H. Durrant-Whyte, and B. Grocholsky. A decentralized architecture for active sensor networks. In 2004 IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04, volume 2, pages 1097–1102 Vol.2. doi: 10.1109/ROBOT.2004.1307971.
- [20] Gonalo S. Martins, David Portugal, and Rui P. Rocha. On the usage of general-purpose compression techniques for the optimization of inter-robot communication. In *Informatics in Control, Automation and Robotics*, pages 223–240. Springer, Cham. URL https://link.springer.com/chapter/10.1007/978-3-319-26453-0_13. DOI: 10.1007/978-3-319-26453-0_13.
- [21] Marius Muja and David G Lowe. Scalable nearest neighbor algorithms for high dimensional data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(11):2227–2240, 2014.
- [22] Eric Nettleton, Sebastian Thrun, Hugh Durrant-Whyte, and Salah Sukkarieh. Decentralised slam with low-bandwidth communication for teams of vehicles. In *Field and Service Robotics*, pages 179–188. Springer, 2003.
- [23] Axel Neumann, Corinna Aichele, Marek Lindner, and Simon Wunderlich. Better approach to mobile ad-hoc networking (batman). *IETF draft, October*, 2008.
- [24] Simon T OCallaghan and Fabio T Ramos. Gaussian process occupancy maps. The International Journal of Robotics Research, 31(1):42–62, 2012.
- [25] Liam Paull, Guoquan Huang, Mae Seto, and John J Leonard. Communicationconstrained multi-auv cooperative slam. In 2015 IEEE International Conference on Robotics and Automation (ICRA), pages 509–516. IEEE, 2015.

- [26] Aleksandr Segal, Dirk Haehnel, and Sebastian Thrun. Generalized-icp. In Robotics: science and systems, volume 2, 2009.
- [27] Jacopo Serafin, Edwin Olson, and Giorgio Grisetti. Fast and robust 3d feature extraction from sparse point clouds. In *Intelligent Robots and Systems (IROS)*, 2016 IEEE/RSJ International Conference on, pages 4105–4112. IEEE, 2016.
- [28] S. Srivastava and N. Michael. Approximate continuous belief distributions for precise autonomous inspection. In 2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), pages 74–80, Oct 2016. doi: 10.1109/SSRR.2016.7784280.
- [29] Bastian Steder, Radu Bogdan Rusu, Kurt Konolige, and Wolfram Burgard. Narf: 3d range image features for object recognition. In Workshop on Defining and Solving Realistic Perception Problems in Personal Robotics at the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), volume 44, 2010.
- [30] James Stephan, Jonathan Fink, and Alejandro Ribeiro. System architectures for communication-aware multi-robot navigation. In Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on, pages 6395–6399. IEEE, 2016.
- [31] Gian Diego Tipaldi and Kai O Arras. Flirt-interest regions for 2d range data. In Robotics and Automation (ICRA), 2010 IEEE International Conference on, pages 3616–3622. IEEE, 2010.
- [32] Yu Zhong. Intrinsic shape signatures: A shape descriptor for 3d object recognition. In Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, pages 689–696. IEEE, 2009.