Bayesian Inference in the Space of Topological Maps

Ananth Ranganathan, Emanuele Menegatti, and Frank Dellaert

***Abstract*— While probabilistic techniques have been considered extensively for performing inference over the space of metric maps, no corresponding general purpose methods exist for topological maps. We present the concept of Probabilistic Topological Maps (PTMs), a sample-based representation that approximates the posterior distribution over topologies given available sensor measurements. The PTM is obtained by performing Bayesian inference over the space of all possible topologies and provides a systematic solution to the correspondence problem in the domain of topological mapping. It is shown that the space of topologies is equivalent to the space of set partitions on the set of available measurements, which is intractably large. This combinatorial problem is overcome by computing an approximate sample-based representation of the posterior. We describe a general framework for modeling measurements and estimating the posterior. A Markov Chain Monte Carlo (MCMC) algorithm that uses specific instances of these models for odometry and appearance measurements is also discussed. We present experimental results that validate our technique and generate good maps when using odometry and appearance as sensor measurements.**

I. Introduction

Mapping an unknown and uninstrumented environment is one of the foremost problems in robotics. Both metric maps [9][33][31] and topological maps [38][3][27][21] have been explored in depth as viable representations of the environment for this purpose. In both cases, probabilistic approaches have had great success in dealing with the inherent uncertainties associated with robot sensori-motor control, that would otherwise make map-building a very brittle process. Lately, the vast majority of probabilistic solutions to the mapping problem also solve the localization problem simultaneously, since these two problems are intimately connected. A solution to the Simultaneous Localization and Mapping (SLAM) problem, as it is called, demands that the algorithm maintain beliefs over the pose of the robot as well as the map of the environment. Subsequently, the pose and the map are each recursively updated by assuming the belief about the other quantity to be fixed and true [31].

The majority of the work in robot mapping deals with the construction of metric maps. Metric maps provide a fine-grained representation of the environment and also contain the actual geometric structure of the environment. This makes navigation using metric maps easy but also introduces significant problems during their construction. Due to systematic errors in odometry, the map tends to drift over time, which makes global consistency difficult to achieve in large environments.

Topological representations, on the other hand, offer a different set of advantages that are useful in many scenarios. Topological maps attempt to capture spatial connectivity of the environment by representing it as a graph with arcs connecting the nodes that are designated significant places in the environment [23]. The arcs are usually annotated with navigation information. This is the definition of a topological map used in our work.

Possibly the hardest problem in robotic mapping is the correspondence problem or data association problem, also variously known as "closing the loop" [15] or "the revisiting problem" [43]. The correspondence problem is the problem of determining if sensor measurements taken at different points in time correspond to the same physical location. When a robot receives a new measurement, it has to decide whether to assign this measurement to one of the locations it has visited previously, or to a completely new location. The correspondence problem is hard as the number of possible choices grows combinatorially. Indeed, we demonstrate below that the number of choices is the same as the number of possible partitions of a set, which grows hyper-exponentially with the cardinality of the set. Previous solutions to the correspondence problem [46][20] commit to a specific correspondence at each step, so that once a wrong decision has been made, the algorithm has difficulty recovering.

In this paper, we describe Probabilistic Topological Maps (PTMs), a sample-based representation that captures the posterior distribution over all possible topological maps given the available sensor measurements. The intuitive reason for computing the posterior is to solve the correspondence problem for topologies in a systematic manner. The set of all possible correspondences between measurements and the physical locations from which the measurements are taken is exactly the set of all possible topologies. By inferring the posterior on this set, it is possible to locate the most probable topologies without committing to a specific correspondence at any point in time, thus providing the most general solution.

The idea of defining a probability distribution over the space of topological maps and using sampling in order to obtain this distribution is the major contribution of this work. The key realization here is that a distribution over this combinatorially large space can be succinctly approximated by a sample set drawn from this distribution. While sampling has been used in the context of data association previously in computer vision [4][5], its use in finding a distribution over all possible maps is completely novel to the best of our knowledge.

As a second major contribution, we show how to perform inference in the space of topologies given uncertain sensor data from the robot, the outcome of which is exactly a Probabilistic Topological Map. A general theory for incorporating odometry and appearance measurements in the inference process is

provided. More specifically, we describe an algorithm that uses Markov chain Monte Carlo (MCMC) sampling [11] to extend the highly successful Bayesian probabilistic framework to the space of topologies. To enable sampling over topologies using MCMC, each topology is encoded as a set partition over the set of landmark measurements. We then sample over the space of set partitions, using as target distribution the posterior probability of the topology given the measurements.

Another important aspect of this work is the definition of a simple but effective prior on the density of landmarks in the environment that we assume. We demonstrate that given this prior the additional sensor information used can be very scant indeed. In fact, while our method is general and can deal with any type of sensor measurement (or, for that matter, prior knowledge), our results include those obtained using only odometry measurements that still yield nice maps of the environment. In addition to using odometry, we also describe an appearance model for use in our algorithm.

Our algorithm is completely data-driven in the sense that it does not require or provide a control algorithm for robot exploration that aids in mapping. Our algorithm also does not compute localization information for the robot during the map inference. Finally, our contribution is not a complete system for topological mapping, but a technique to compute a posterior distribution over topologies given the sensor measurements from the landmark locations. Accordingly, we do not provide landmark detection algorithms or other techniques for detecting "significant places", but assume that these are available.

In subsequent sections, we first provide related work in probabilistic mapping in general and topological mapping in particular. Then, we define Probabilistic Topological Maps formally and provide the theory for estimating the posterior over the space of topologies. Subsequently, we describe an implementation of the theory using MCMC sampling in topological space. This is followed by a section that provides details about the specific odometry and appearance models used and their evaluation. In particular, we use Fourier Signatures [17][28] of panoramic images to construct an appearance model in this case. A prior over the space of topologies is also described. Finally, we provide experimental validation for our technique and conclude the paper with a discussion about our method.

II. Related Work

Our work is drawn from the area of probabilistic mapping and, more specifically, topological mapping. We review prior research in these areas that is relevant to our work.

A. Probabilistic Mapping and SLAM

Early approaches to the mapping problem (usually obtained by solving the SLAM problem) used Kalman filters and Extended Kalman filters [24][2][6][8][41][42]. Kalman filter approaches assume that the motion model, the perceptual model (or the measurement model) and the initial state distribution are all Gaussian. Extended Kalman filters relax these assumptions a bit by linearizing the motion model using a Taylor series expansion. More importantly, the Kalman filter approach can estimate the complete posterior over maps efficiently. This is offset by their inability to cope with the correspondence problem.

A well-known extension of the basic Kalman filter paradigm is the Lu/Milios algorithm [26], a laser-specific algorithm that performs maximum likelihood correspondence. It iterates over a map estimation and a data association phase that enable it to recover from wrong correspondences in the presence of small errors. In spite of this, the algorithm encounters limitations when faced with large pose errors and fails in large environments.

Rao-Blackwellized Particle Filters (RBPFs) [34], of which the FastSLAM [31][32] algorithm is a specific implementation, are also theoretically capable of maintaining the complete posterior distribution over maps. This is possible since each sample in the RBPF can represent a different data association decision [30]. However, in practice the dimensionality of the trajectory space is too large to be adequately represented in this approach, and often the ground-truth trajectory along with the correct data association will be missed altogether. This problem is a fundamental shortcoming of the importance sampling scheme used in the RBPF, and cannot be dealt with satisfactorily except by an exponential increase in the number of samples, which is intractable. Additionally, RBPFs are also prone to odometry drift over time. Recent work by Haehnel et. al. [15] tries to overcome the odometry drift by correcting for it through scan matching.

Yet another approach to SLAM that has been successful is the use of the EM algorithm to solve the correspondence problem in mapping [46][1]. The algorithm iterates between finding the most likely robot pose and the most likely map. EM-based algorithms do not compute the complete posterior over maps, but instead perform hill-climbing to find the most likely map. Such algorithms make multiple passes over sensor data which makes them extremely slow and unfit for on-line, incremental computation. In addition, EM cannot overcome local minima, resulting in incorrect data associations. Other approaches exist that report loop closures and re-distribute the error over the trajectory [14][35][45][43], but these decisions are again irrevocable and hence mistakes cannot be corrected.

Recent work by Duckett [7] on the SLAM problem is similar to our own, in the sense that he too searches over the space of possible maps. The SLAM problem is presented as a global optimization problem and metric maps are coded as chromosomes for use in a genetic algorithm. The genetic algorithm searches over the space of maps (or chromosomes) and finds the most likely map using a fitness function. However, since genetic algorithms are susceptible to local minima, this method suffers from the same shortcomings as the EM techniques described previously.

B. Topological Maps

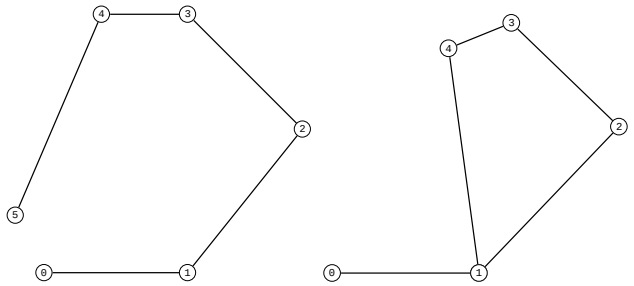
Many topological approaches to mapping include robot control to help maneuver the robot to the exact location it was in when visiting the location previously or to guide the robot around a suspected loop again. This helps solve the correspondence problem. Examples of this approach include Choset's

Generalized Voronoi Graphs [3] and Kuipers' Spatial Semantic Hierarchy [23]. Other approaches that involve behavior-based control for exploration-based topological mapping are also fairly common. Mataric [27] uses boundary-following and goal-directed navigation behaviors in combination with qualitative landmark identification to find a topological map of the environment. A complete behavior-based learning system based on the Spatial Semantic Hierarchy that learns at many levels starting from low-level sensori-motor control to topological and metric maps is described in [37]. Yamauchi et al. [50][51] use a reactive controller in conjunction with an Adaptive Place Network that detects and identifies special places in the environment. These locations are subsequently placed in a network denoting spatial adjacency.

Though probabilistic methods have been used in conjunction with topological maps before, none exist that are capable of dealing with the inference of a posterior distribution in topological space. Most instances of previous work extant in the literature that incorporate uncertainty in topological map representations do not deal with general topological maps, but with the use of decision theory to learn a policy that the robot follows to navigate the environment. Simmons and Koenig [40] model the environment using a POMDP in which observations are used to update belief states. Shatkay and Kaelbling [39] use the Baum-Welch algorithm, a variant of the EM algorithm used in the context of HMMs, to solve the correspondence problem for topological mapping. Other examples of such work include [19] and [13].

Lisien et al. [25] have provided a method that combines locally estimated feature-based maps with a global topological map. Data association for the local maps is performed using a simple heuristic wherein each measurement is associated with the existing landmark having the minimum distance to the measured location. A new landmark is created if this distance is above a threshold. The set of local maps is then combined using an "edge-map" association, i.e. the individual landmarks are aligned and the edges compared. Clearly, this technique, while suitable for mapping environments where the landmark locations are sufficiently dissimilar, is not robust in environments with large or multiple loops.

Another approach that is closer to the one presented here is given by Tomatis et al. [49] and uses POMDPs to solve the correspondence problem. However, in their case while a multi-hypothesis space is maintained, it is used only to detect the points where the probability mass splits into two. Also, this work like a lot of others uses specific qualities of the indoor environment such as doors and corridor junctions, and hence is not generally applicable to any environment. Other work by Kuipers and Beeson [22] focusses on the identification of distinctive places which is accomplished by applying a clustering algorithm to the measurements at all the distinctive places. Unlike our method, this method does not include inference about the topologies themselves. Finally, SLAM algorithms used to generate metric maps have also been applied to generating integrated metric and topological maps with some success. For instance, Thrun et al. [47] use the EM algorithm to solve the correspondence problem while building a topological map. The computed correspondence is subsequently used in constructing a metric map. By contrast, Thrun [44] first computes a metric map using value iteration and uses thresholding and Voronoi diagrams to extract the topology from this.



(a) (b)

Fig. 1. Two topologies with 6 observations each corresponding to set partitions (a) with six landmarks $(\{0\},\{1\},\{2\},\{3\},\{4\},\{5\})$ and (b) with five landmarks$(\{0\},\{1,5\},\{2\},\{3\},\{4\})$ where the second and sixth measurement are from the same landmark.

III. Probabilistic Topological Maps

A Probabilistic Topological Map is a sample-based representation that approximates the posterior distribution $P(T|Z)$ over topologies T given observations Z. While the space of possible maps is combinatorially large, a probability density over this space can be approximated by drawing a sample of possible maps from the distribution. Using the samples, it is possible to construct a histogram on the support of this sample set.

For the purpose of this work, we assume that the robot is equipped with a "landmark detector" that simply recognizes a landmark when it is near (or on) a landmark, i.e. it is a binary measurement that tells us when landmarks were encountered. No knowledge of the correspondence between landmark observations and the actual landmarks is given to the robot: indeed, that is exactly the topology that we seek. The problem then is to compute the discrete posterior probability distribution $P(T|Z)$ over the space of topologies.

Our technique exploits the equivalence between topologies of an environment and set partitions of landmark measurements, which group the measurements into a set of equivalence classes. When all the measurements of the same landmark are grouped together, this naturally defines a partition on the set of measurements. It can be seen that a topology is nothing but the assignment of measurements to sets in the partition, resulting in the above mentioned isomorphism between topologies and set partitions. An example of the encoding of topologies as set partitions is shown in Figure 1.

We begin our consideration by assuming that the robot observes N "special places" or landmarks during a run, not all of them necessarily distinct. Formally, for the N element measurement set $Z = \{Z_i | 1 \leq i \leq N\}$, a partition T can be represented as $T = \{S_i \mid i \in [1, M]\}$, where each S_i is a set of measurements such that $S_i \cap S_j = \phi$ $\forall i, j \in [1, M]$, $i \neq j$, $\bigcup_{i=1}^{M} S_i = Z$, and $M \leq N$ is the number of sets in the partition. M is also the number

of distinct landmarks in the environment. In the context of topological mapping, all members of the set S_i represent landmark observations of the ith landmark. The cardinality of the set of all possible topologies is identical to the number of set partitions of the observation N-set. This number is called the Bell number b_N [36], defined as $b_N = \frac{1}{e}\sum_{k=0}^{\infty}\frac{k^N}{k!}$, and grows hyper-exponentially (but slower than the factorial) with N, for example $b_1 = 1$, $b_2 = 5$ but b_{15} =190899322. The combinatorial nature of this space makes exhaustive evaluation impossible for all but trivial environments.

IV. A General Framework for Inferring PTMs

The aim of inference in the space of topologies is to obtain the posterior probability distribution on topologies $P(T|Z)$. All inference procedures that compute sample-based representations of distributions require that evaluation of the sampled distribution be possible. In this section, we describe the general theory for evaluating the posterior at any given topology.

Using Bayes Law on the posterior $P(T|Z)$, we obtain

$$P(T|Z) \propto P(Z|T)P(T) \tag{1}$$

where $P(T)$ is a prior on topologies and $P(Z|T)$ is the observation likelihood.

In this work, we assume that the only observations we possess are odometry and appearance. Note that this is not a limitation of the framework, and other sensor measurements, such as laser range scans, can easily be taken into consideration. We factor the set Z as $Z = \{O, A\}$, where O and A correspond to the set of odometry and appearance measurements respectively. This allows us to rewrite (1) as

$$\begin{aligned} P(T \mid O, A) &= kP(O, A|T)P(T) \\ &= kP(O|T)P(A|T)P(T) \end{aligned} \tag{2}$$

where k is the normalization constant, and we have used the fact that the appearance and odometry are conditionally independent given the topology. We discuss evaluation of the appearance likelihood $P(A|T)$, odometry likelihood $P(O|T)$, and the prior on topologies $P(T)$, in the following sections.

A. Evaluating the Odometry Likelihood

It is not possible to evaluate the odometry likelihood $P(O|T)$ without knowledge of the landmark locations. However, since we are not interested in the landmark locations, we integrate over the set of landmark locations X and calculate the marginal distribution $P(O|T)$ from the joint distribution $P(O, X|T)$. The likelihood is then the following integral:

$$P(O|T) = \int_X P(O|X, T)P(X|T) \tag{3}$$

where $P(O|X, T)$ is the measurement model, an unknown density on O given X and T, and $P(X|T)$ is a prior over landmark locations. Note that (3) makes no assumptions about the actual form of X, and hence, is completely general. Evaluation of the odometry likelihood using (3) requires the specification of a prior distribution $P(X|T)$ over landmark locations in the environment and a measurement model $P(O|X, T)$ for the odometry given the landmark locations.

B. Evaluating the Appearance Likelihood

Similar to the estimation of the odometry likelihood, which was performed by introducing the set of landmark locations, estimation of the appearance likelihood $P(A \mid T)$ is performed through the introduction of a hidden parameter $Y = \{y_i|1 \leq i \leq M\}$ which denotes the "*true appearance*" corresponding to each set in the topology. As we do not care about computing this value we marginalize over Y , so that

$$P(A \mid T) = \int_Y P(A \mid Y, T)P(Y \mid T) \tag{4}$$

where $P(Y \mid T)$ is a prior on the appearance. As each individual y_i is independent of all other y_j, the prior $P(Y \mid T)$ can be factored into a product of priors on the individual y_i.

$$P(Y \mid T) = \prod_{i=1}^{M} P(y_i) \tag{5}$$

The topology T introduces a partition on the set of appearance measurements by determining which "true appearance" y_i each measurement a_{ij} actually measures, i.e the partition encodes the correspondence between the set a and the set y. Also, given Y, the likelihood of the appearance can be factored into a product of likelihoods of the individual appearance instances. Hence, denoting the ith set in the partition as S_i, we rewrite $P(A \mid Y, T)$ as -

$$P(A \mid Y, T) = \prod_{i=1}^{M}\prod_{j=1}^{|S_i|} P(a_{ij} \mid y_i) \tag{6}$$

where the dependence on T is subsumed in the partition. Combining Equations (4), (5) and (6), we get the expression for the appearance likelihood as

$$P(A \mid T) = \prod_{i=1}^{M}\int_{y_i} P(y_i)\prod_{j=1}^{|S_i|} P(a_{ij} \mid y_i) \tag{7}$$

In the above equation, $P(y_i)$ is a prior on appearance in the environment, and $P(a_{ij} \mid y_i)$ is the appearance measurement model. Evaluation of the appearance likelihood requires the specification of these two quantities.

C. Prior on Topologies

The prior on topologies $P(T)$, required to evaluate (2), assigns a probability to topology T based on the number of distinct landmarks in T and the total number of measurements. The prior is easily obtained and intuitively understood using an urn-ball model.

Let the total number of landmarks in the environment be L, and let N and M be the number of measurements and number of landmarks in the topology as before. It is possible that L is greater than N. This setup can be converted into an urn-ball model by considering landmarks to be urns and measurements to be balls, giving L urns and N balls respectively. A set partition on the measurements is created by randomly adding the balls to the urns, where it is assumed that a ball is equally likely to land in any urn (i.e. there is a uniform distribution on the urns). The distribution on the number of occupied urns,

after adding all the N balls randomly to the urns, is given by the Classical Occupancy Distribution [18] as

$$P(M) = \binom{L}{M} L^{-N} M! \left\{ {N \atop M} \right\} \quad (8)$$

where $\left\{ {N \atop M} \right\}$ denotes the Stirling number of the second kind that gives the number of possible ways to split a set of size N into M subsets, and is defined recursively as $\left\{ {N \atop M} \right\} \triangleq \left\{ {N-1 \atop M-1} \right\} + M\left\{ {N-1 \atop M} \right\}$ [36].

The number of occupied urns after adding all the balls corresponds to the number of landmarks in the topology, while the specific allocation of balls to urns (called an allocation vector) corresponds to the topology itself. Also, (8) assigns an equal probability to all ball allocations with the same number of occupied urns. Hence, we can interpret (8) as

$$P(M) \propto P(\text{allocation vector with } M \text{ occupied urns}) \times \text{No. of allocation vectors with } M \text{ occupied urns} \quad (9)$$

The number of allocation vectors with M occupied urns is equal to the number of partitions of the set of balls into M subsets. This is precisely the Stirling number of the second kind $\left\{ {N \atop M} \right\}$. Combining this observation with (8) and (9) yields

$$P(\text{allocation with } M \text{ occupied urns}) \propto \binom{L}{M} L^{-N} M! = \frac{L!}{(L-M)!} L^{-N}$$

As mentioned previously, the probability of an allocation vector corresponds to the probability of a topology. Hence, the prior probability of a topology T with M landmarks is

$$P(T|L) = k \frac{L^{-N} \times L!}{(L-M)!} \quad (10)$$

where k is a normalization constant. This prior distribution assigns equal probability to all topologies containing the same number of landmarks.

Note that the total number of landmarks L is not known. Hence, we assume a Poisson prior on L, giving $P(L|\lambda) = \frac{\lambda^L e^{-\lambda}}{L!}$, and marginalize out L to get

$$\begin{aligned} P(T) &= \sum_L P(T|L) P(L|\lambda) \\ &\propto \sum_L \frac{L^{-N} \times \lambda^L e^{-\lambda}}{(L-M)!} \\ &= e^{-\lambda} \sum_{L=M}^{\infty} \frac{L^{-N} \times \lambda^L}{(L-M)!} \end{aligned} \quad (11)$$

where λ is the Poisson parameter and the summation replaces the integral as the Poisson distribution is discrete. In practice, the prior on L is actually a truncated Poisson distribution since the summation in (11) is only evaluated for a finite number of terms. Specifying a different distribution on the allocation of balls to urns, rather than the uniform distribution assumed above, yields different priors on topologies. We do not further explore this possibility in this work.

Algorithm 1 The Metropolis-Hastings algorithm

1) Start with a valid initial topology T_t, then iterate once for each desired sample
2) Propose a new topology $T_t^{'}$ using the *proposal distribution* $Q(T_t'; T_t)$
3) Calculate the *acceptance ratio*

$$a = \frac{P(T_t'|Z^t)}{P(T_t|Z^t)} \frac{Q(T_t; T_t')}{Q(T_t'; T_t)} \quad (12)$$

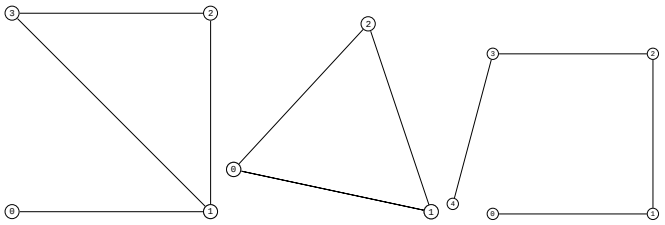
where $Z^t$is the set of measurements observed up to and including time t.
4) With probability $p = \min(1, a)$, accept T_t' and set $T_t \leftarrow T_t'$. If rejected we keep the state unchanged (i.e. return T_t as a sample).

V. Inferring Probabilistic Topological Maps using MCMC

The previous section provided a general theory for inferring the posterior over topologies using odometry and appearance information. We now present a concrete implementation of the theory that uses the Metropolis-Hastings algorithm [16], a very general MCMC method, for performing the inference. All MCMC methods work by running a Markov chain over the state space with the property that the chain ultimately converges to the target distribution of our interest. Once the chain has converged, subsequent states visited by the chain are considered to be samples from the target distribution. The Markov chain itself is generated using a proposal distribution that is used to propose the next state in the chain, a move in state space, possibly by conditioning on the current state. The Metropolis-Hastings algorithm provides a technique whereby the Markov chain can converge to the target distribution using any arbitrary proposal distribution, the only important restriction being that the chain be capable of reaching all the states in the state space.

The pseudo-code to generate a sequence of samples from the posterior distribution $P(T|Z)$ over topologies T using the Metropolis-Hastings algorithm is shown in Algorithm 1 (adapted from [11]). In this case the state space is the space of all set partitions, where each set partition represents a different topology of the environment. Intuitively, the algorithm samples from the desired probability distribution $P(T|Z)$ by rejecting a fraction of the moves generated by a proposal distribution $Q(T_t'; T_t)$, where T_t is the current state and T_t' is the proposed state. The fraction of moves rejected is governed by the acceptance ratio a given by (12), which is where most of the computation takes place. Computing the acceptance ratio, and hence, sampling using MCMC, requires the design of a proposal density and evaluation of the target density, the details of which are discussed below.

We use a simple split-merge proposal distribution that operates by proposing one of two moves, a split or a merge with equal probability at each step. Given that the current sample topology has M distinct landmarks, the next sample is obtained by splitting a set, to obtain a topology with $M+1$ landmarks, or merging two sets, to obtain a topology with $M-1$ landmarks. If the chosen move is not possible, for example a merge move may not be possible because the



(a) ({0},{1,4},{2},{3}) (b) ({0,3},{1},{4},{2}) (c) ({0},{1},{2},{3},{4})

Fig. 2. Illustration of the proposal - Given a topology (a) corresponding to the set partition with N=5, M=4, the proposal distribution can (b) perform a merge step to propose a topology with a smaller number of landmarks corresponding to a set partition with N=5, M=3 or (c) perform a split step to propose a topology with a greater number of landmarks corresponding to a set partition with N=M=5 or re-propose the same topology.

topology only contains one landmark, the current topology is re-proposed. The proposal is illustrated in Figure 2 for a trivial environment.

The *merge move* merges two randomly selected sets in the partition to produce a new partition with one less set than before. The probability of a merge is simply $1/N_M$ where N_M is the number of possible merges and is equal to the binomial coefficient $\binom{M}{2}$, ($M > 1$).

The *split move* splits a randomly selected set in the partition to produce a new partition with one more set than before. To calculate the probability of a split move, let N_S be the number of non-singleton sets in the partition. Clearly, N_S is the number of sets in the partition that can be split. Out of these N_S sets, we pick a random set R to split. The number of possible ways to split R into two subsets is given by the Stirling number $\left\{ {|R| \atop 2} \right\}$. Combining the probability of selecting R and the probability of splitting it, we obtain the probability of the split move as $p_{\text{split}} = \left(N_S \left\{ {|R| \atop 2} \right\} \right)^{-1}$.

The proposal distribution is summarized in pseudo-code format in Algorithm 2, where q is the proposal distribution and $r = \frac{q(T'|T)}{q(T|T')}$ is the proposal ratio, a part of the acceptance ratio in Algorithm 1. It is to be noted that this proposal does not incorporate any domain knowledge, but uses only the combinatorial properties of set partitions to propose random moves

In addition to proposing new moves in the space of topologies, we also need to evaluate the posterior probability $P(T|Z)$. This is done as described in Section IV. The specification of the measurement models and the details of evaluating the posterior probability using these models are given in the following section.

VI. Evaluating the Posterior Distribution

We evaluate the posterior distribution, which is also the MCMC target distribution, using (2). It is important to note that we do not need to calculate the normalization constant in (2) since the Metropolis-Hastings algorithm requires only a ratio of the target distribution evaluated at two points, wherein the normalization constant cancels out. The odometry and

Algorithm 2 The Proposal Distribution

1) Select a merge or a split with probability 0.5
2) If a merge is selected go to 3, else go to 4
3) **Merge move:**
   - if T contains only one set, re-propose $T' = T$, hence $r = 1$
   - otherwise select two sets at random, say P and Q
     a) $T' = (T - \{P\} - \{Q\}) \cup \{P \cup Q\}$ and $q(T'|T) = \frac{1}{N_M}$
     b) $q(T|T')$ is obtained from the reverse case 4(b), hence $r = N_M^{-1} N_S \left\{ {|P \bigcup Q| \atop 2} \right\}$, where N_S is the number of possible splits in T'
4) **Split move:**
   - if T contains only singleton sets, re-propose $T' = T$, hence $r = 1$
   - otherwise select a non-singleton set R at random from T and split it into two sets P and Q.
     a) $T' = (T - \{R\}) \cup \{P, Q\}$ and $q(T'|T) = \left(N_S \left\{ {|R| \atop 2} \right\} \right)^{-1}$
     b) $q(T|T')$ is obtained from the reverse case 3(b), hence $r = N_M \left(N_S \left\{ {|R| \atop 2} \right\} \right)^{-1}$, where N_M is the number of possible merges in T'

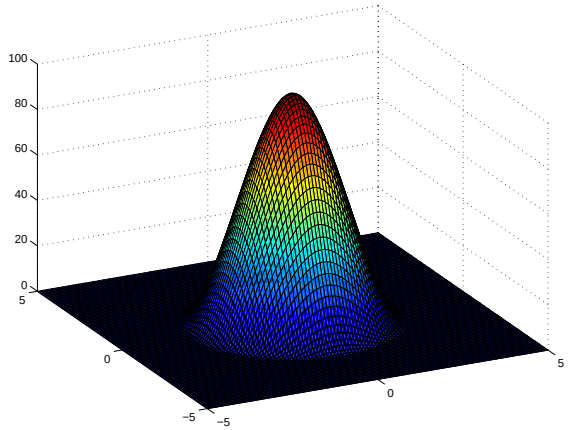


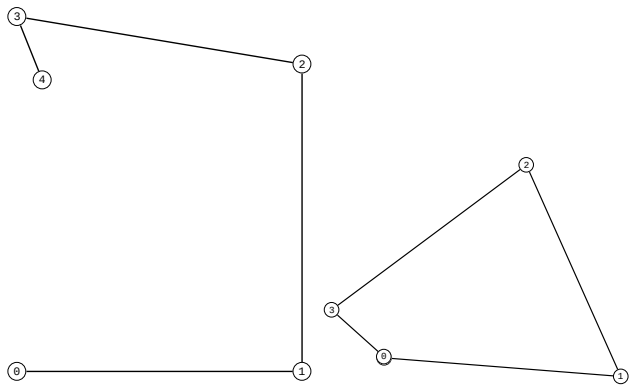
Fig. 3. Cubic penalty function (in this case, with a threshold distance of 3 meters) used in the prior over landmark density

appearance measurement models required to evaluate (2) are described below.

A. Evaluating the Odometry Likelihood

Evaluation of the odometry likelihood is performed using (3) under the assumption, common in robotics literature, that landmark locations have the 2D form $X = \{(x_i, y_i, \theta_i) | 1 \leq i \leq N\}$. This requires the definition of a prior on the distribution of the landmark locations X conditioned on the topology T, $P(X|T)$.

We use a simple prior on landmarks that encodes our assumption that landmarks do not exist close together in the environment. If the topology T places two distinct landmarks x_i and x_j within a distance d of each other, the negative log likelihood corresponding to the two landmarks is given by the



(a) (b)

Fig. 4. Illustration of the optimization using "soft constraints". The observed odometry in (a) is modified to the one in (b) because the soft constraint determined by the topology used in this case, $(\{0,4\},\{1\},\{2\},\{3\})$, tries to place the first and last landmarks at the same physical location.

penalty function

$$L(x_i, x_j; T) = L(x_j, x_i; T) \quad = \quad \begin{cases} f(d) & d < D \\ 0 & d \geq D \end{cases} \quad (13)$$

where d is the Euclidean distance between x_i and x_j, D is a threshold value, called the "penalty radius", and we define $f(d)$ to be a cubic function as shown in Figure 3. The cubic function is defined using two parameters - the penalty radius D at which the function becomes zero, and the maximum value of the function at the origin. The total probability of landmark locations X given topology T, $P(X|T)$, is then calculated as

$$P(X|T) = \exp\left\{ - \sum_{\substack{1 \leq i < j \leq N \\ x_j \notin S(x_i)}} L\left(x_i, x_j\right) \right\} \quad (14)$$

where $S(x_i)$ denotes the set containing x_i.

Now, we can write the negative log-likelihood function corresponding to $P(O|T)$ in (3) as

$$L_O(X) = \left(\frac{X - X_O}{\sigma_O}\right)^2 + \sum_{S \in T} \sum_{i,j \in S} \left(\frac{X_i - X_j}{\sigma_T}\right)^2 + \sum_{\substack{1 \leq i < j \leq N \\ x_j \notin S(x_i)}} L\left(x_i, x_j\right) \quad (15)$$

where S is a set in the partition corresponding to T, σ_O and σ_T are standard deviations explained below, and X_o is the set of landmark locations obtained from the odometry measurements. The intuition here is that the topology T constrains some measurements as being from the same location even though the odometry may put these locations far apart. The log-likelihood function accounts for the error from distorting the odometry, the first term in (16), and the error for not conforming to the topology T, the second term in (16).

The error for not conforming to a topology is expressed through a set of "soft constraints". These constraints try to place two observations that are ascribed to the same landmark by the topology at the same physical location. A simple example illustrating soft constraints is given in Figure 4. The standard deviations for the odometry and soft constraints, σ_O and σ_T respectively, encode the amount of error that we are willing to tolerate in each of these quantities. The final term in (15), where the sum is over all X_i and X_j not contained in the same set, is simply the negative log-likelihood of the prior in (14).

1) Numerical Evaluation of the Odometry Likelihood: In some cases, it may be possible to evaluate the integral in (3) analytically using the functional form of the log-likelihood given in (15). If closed form evaluation is not possible, it may still be possible to use an analytical approximation technique such as Laplace's method [48] to evaluate (3).

However, in general, it is not possible to use any form of analytical evaluation to compute (3). Instead, we employ a Monte Carlo approximation, using importance sampling [10] to approximate the integrand $P(O|X,T)P(X|T)$. Importance sampling works by generating samples from a proposal distribution that is easy to sample from. Each sample is then weighted by the ratio of the target distribution to the proposal distribution evaluated at the sample location. The Monte Carlo approximation is subsequently performed by summing the weighted samples. The primary condition on the proposal distribution is that it should be non-zero at all locations where the target distribution is non-zero. In addition, importance sampling is efficient if the proposal distribution is a close approximation to the target distribution.

In our case, the importance sampling proposal distribution is obtained by modifying the log-likelihood in (15). Firstly, ignoring the final term corresponding to the prior in (15), we obtain the function

$$\psi(X) \quad = \quad \left(\frac{X - X_O}{\sigma_O}\right)^2 + \sum_{S \in T} \sum_{i,j \in S} \left(\frac{X_i - X_j}{\sigma_T}\right)^2 \quad (16)$$

This function is a lower bound on (15) since the prior term in (15) is never negative. Consequently, (16) can be used to obtain a valid importance sampling distribution. We then employ Laplace's method to obtain a multivariate Gaussian distribution from $\psi(X)$, which is used as the proposal distribution. This is achieved by computing the maximum likelihood path $X^\star$ through a non-linear optimization of $\psi(X)$, and creating a local Gaussian approximation $Q(X|O,T)$ around $X^\star$

$$X^\star = \underset{X}{\text{argmax}} \;\; \psi(X)$$

$$Q\left(X \mid O, T\right) = \frac{1}{\sqrt{|2\pi\Sigma|}} e^{-\frac{1}{2}(X - X^\star)^T \Sigma^{-1} (X - X^\star)}$$

where Σ is the covariance matrix relating to the curvature of $\psi(X)$ around X^*. The distribution $Q(X|O,T)$ is then used as the proposal distribution for the importance sampler.

In practice, we use the Levenberg-Marquardt algorithm in conjunction with a sparse QR solver to perform the optimization described above. The Levenberg-Marquardt algorithm requires the derivative of the objective function that is being minimized, in this case the function $\psi(X)$ in (16). To

compute the (sparse) Jacobian H given by $H = \frac{\partial \psi(X)}{\partial X}$, we use an automatic differentiation (AD) framework. Automatic differentiation (AD) is a technique for augmenting computer programs with derivative computations. It exploits the fact that by applying the chain rule of differential calculus repeatedly to elementary operations, derivatives of arbitrary order can be computed automatically and accurately to working precision. See [12] for more details.

The odometry likelihood given by (3) is now evaluated using the Monte Carlo approximation

$$\int_X P(O|X,T)P(X|T) \approx \frac{1}{N}\sum_{i=1}^{N} \frac{P(O|X^{(i)},T)P(X^{(i)}|T)}{Q(X^{(i)}|O,T)} \tag{17}$$

where the $X^{(i)}$ are samples obtained from the Gaussian proposal distribution $Q(X|O,T)$ and N is the number of samples.

B. Evaluating the Appearance Likelihood

We use Fourier signatures as appearance measurements. Fourier signatures have previously been used in the context of memory-based navigation [28] and localization using omni-directional vision [29]. They are computed by calculating the 1-D Fourier transform of each row of the panoramic image and storing these coefficients in a matrix [29]. Only a few coefficients corresponding to the lower spatial frequencies are used for this purpose.

A more popular dimensionality reduction technique is to extract a set of eigen images from the set of measurement images and project the images onto this eigen space. The drawback of such systems is that they need to further pre-process the measurement images in order to obtain rotational invariance. In contrast, the Fourier coefficient magnitudes in a Fourier signature are rotation-invariant as panoramic images are periodic. Hence, a Fourier signature yields a low-dimensional, rotation-invariant representation of the image. We use images obtained from an eight-camera rig mounted on a robot to produce panoramic images. The eight images obtained at each point in time are stitched together automatically to form a 360^0 view of the environment. An example of such a panoramic image is shown in Figure 5.

In our case, Fourier signatures are calculated using a modification of the procedure given in [28]. Firstly, a single column image obtained by averaging the columns of the input image is calculated and subsequently, the one-dimensional Fourier transform of this image is performed. This gives us the Fourier signature of the image. It is to be noted that Fourier signatures do not comprise a robust source of measurements, since the measurements contain many false positives, in the sense that images from distinct physical locations often yield similar Fourier signatures. However, they have the advantage of being simple to compute and model. Moreover, in conjunction with odometry, they still produce good results as we demonstrate in Section VII.

Evaluation of the appearance likelihood is performed using (7). However, in this case, each appearance measurement a_{ij} is a vector given as $a_{ij} = \{a_{ij1}, a_{ij2}, \ldots, a_{ijK}\}$, where a_{ijk} is the kth Fourier component in the Fourier signature. Also, we assume a similar vector form for the hidden appearance variables y_i, so that $y_i = \{y_{i1}, y_{i2}, \ldots, y_{iK}\}$. We can then write (7) as

$$P(A|T) = \prod_{i=1}^{M} \int_{y_i} P(y_{i1}, \ldots, y_{iK}) \times \prod_{j=1}^{|S_i|} P(a_{ij1}, \ldots, a_{ijK} \mid y_{i1}, \ldots, y_{iK}) \tag{18}$$

Clearly, the various frequency components of the Fourier signature are independent given the corresponding appearance variable, and hence, can be factored, as can be the prior over the hidden appearance variables. Consequently, we modify (18) to get the expression for the appearance likelihood as

$$P(A \mid T) = \prod_{i=1}^{M}\prod_{k=1}^{K} \int_{y_{ik}} P(y_{ik}) \prod_{j=1}^{|S_i|} P(a_{ijk} \mid y_{ik}) \tag{19}$$

We assume the measurement model for appearance instance a_{ijk} to be Gaussian centered around y_{ik} with variance σ_{ik}^2. Hierarchical priors are placed on y_{ik} and σ_{ik}^2: the prior on σ_{ik}^2 being an inverse gamma distribution while the prior on y_{ik} is taken to be a Gaussian with mean μ and variance $\frac{\sigma_{ik}^2}{\kappa}$. This particular choice of priors forms the conjugate prior to the Gaussian measurement model with unknown mean and variance, and allows the integration in (19) to be performed analytically. The appearance model can then be summarized as

$$\begin{aligned} a_{ijk} &\sim \mathcal{N}(y_{ik}, \sigma_{ik}^2) \\ y_{ik} &\sim \mathcal{N}(\mu, \frac{\sigma_{ik}^2}{\kappa}) \\ \sigma_{ik}^2 &\sim IG(\alpha_k, \beta_k) \end{aligned} \tag{20}$$

where IG denotes the inverse gamma distribution. Note that while the value of κ is generally chosen so that the prior on y_{ik} is vague, we usually have some extra "world knowledge" that can be used to set the values of the hyper-parameters α_k and β_k. For example, if we expect the value of the Fourier signature to vary by only a small amount in the neighborhood of a given location, the prior on σ_{ik}^2 should reflect this knowledge by being peaked about a specific value.

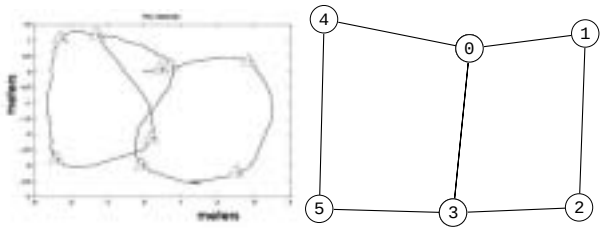
The generative model for Fourier signature measurements specified by (20) is now used to compute the appearance likelihood given by (19) as follows

$$P(A \mid T) = \prod_{i=1}^{M}\prod_{k=1}^{K} \int_{\sigma_{ik}^2} IG(\alpha_k, \beta_k) \times \int_{y_{ik}} \mathcal{N}(\mu, \frac{\sigma_{ik}^2}{\kappa}) \prod_{j=1}^{|S_i|} \mathcal{N}(y_{ik}, \sigma_{ik}^2) \tag{21}$$

Due to the use of the conjugate prior, we can compute $P(A|T)$ from (21) analytically. We prove in the appendix that performing the integration over y_{ik} and σ_{ik}^2 gives the expression for the appearance likelihood as

$$P(A|T) \propto \prod_{i=1}^{M} C_i^K \prod_{k=1}^{K} \Gamma(\gamma_{ik}+1) \left(\beta + \frac{1}{2}\Phi_{ik}\right)^{-(\gamma_i+1)} \tag{22}$$

Fig. 5. A Panoramic image obtained from the robot camera rig



(a) (b)

Fig. 6. (a) Raw odometry and (b) Ground truth topology from the first experiment involving 9 observations

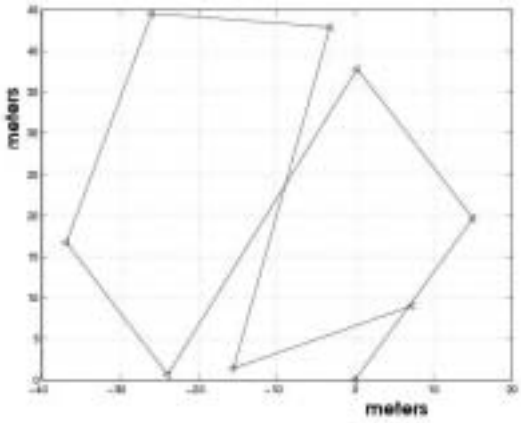


Fig. 7. Landmark locations plotted using odometry for second experiment

where

$$
\begin{aligned}
C_i &= (\kappa + |S_i|)^{-\frac{1}{2}} \\
\Phi_{ik} &= \kappa\,(\mu^\star - \mu)^2 + \sum_{j=1}^{|S_i|} (a_{ijk} - \mu^\star)^2 \\
\mu^\star &= \frac{\kappa\mu + \sum_{j=1}^{|S_i|} a_{ijk}}{\kappa + |S_i|} \\
\gamma_i &= \alpha + \frac{|S_i|}{2} + 1
\end{aligned}
$$

and constants that do not affect the likelihood ratio have been omitted.

The appearance model presented above is not specific to Fourier signatures. Indeed, it is a general purpose clustering model that assumes that the data to be clustered are distributed as a mixture of Gaussians with an unknown number of components. A topology labels each data instance as arising out of one of the mixture components, where the number of mixture components is determined by the topology.

C. Putting it Together

The odometry and appearance likelihoods and the prior on topologies required to compute the target distribution (2), are given by (17), (22) and (11) respectively. We use this target distribution to sample using Algorithm 1 as explained before.

VII. Experiments and Results

Two sets of experiments were performed to validate the Probabilistic Topological Maps algorithm. Both the experiments were performed using an ATRV-Mini mounted with an eight-camera rig. In all cases, we initialized the sampler with the partition that assigned each measurement to its own set. We describe the experiments and results below.

The first experiment was conducted inside our lab using a relatively short run of the robot. Nine landmark locations were observed during the run of approximately 15 meters. The raw odometry obtained from the robot, labeled with the landmark locations, and the ground-truth topology are shown in Figure 6. Only the odometry measurements were used in the experiment, no appearance information was provided to the algorithm. This was done by simply neglecting the appearance likelihood term in (2). The penalty radius was set to 2.5 meters for this experiment.

Table I shows the evolution of the MCMC sampler for different values of the maximum penalty parameter. In our algorithm, it is the penalty term that facilitates merging of nodes in the map that are the same. Without the penalty, the system has no incentive to move toward a topology with lesser number of nodes as this increases the odometry error. Table I(a) illustrates this case. It can be seen that the topology that is closest to the raw odometry data and also having the maximum possible nodes gets the maximum probability mass. For the rest of the cases with maximum penalties equal to 50, 100, and 150 respectively, the most likely solution is the topology indicated by the raw odometry. The large error in odometry makes the ground truth topology less likely in these cases. The ground truth topology is the second-most likely topology for maximum penalty values 100 and 150. This is because as the penalty is increased the effect of odometry is diminished and the ground truth topology gains more of the probability mass. However, a very large penalty swamps odometry data and makes absurd topologies more likely.

The second experiment demonstrates the usefulness of appearance in disambiguating noisy odometry measurements. The experiment was conducted in an indoor office environment where the robot traveled along the corridors in a run of approximately 200 meters and observed nine landmarks. The

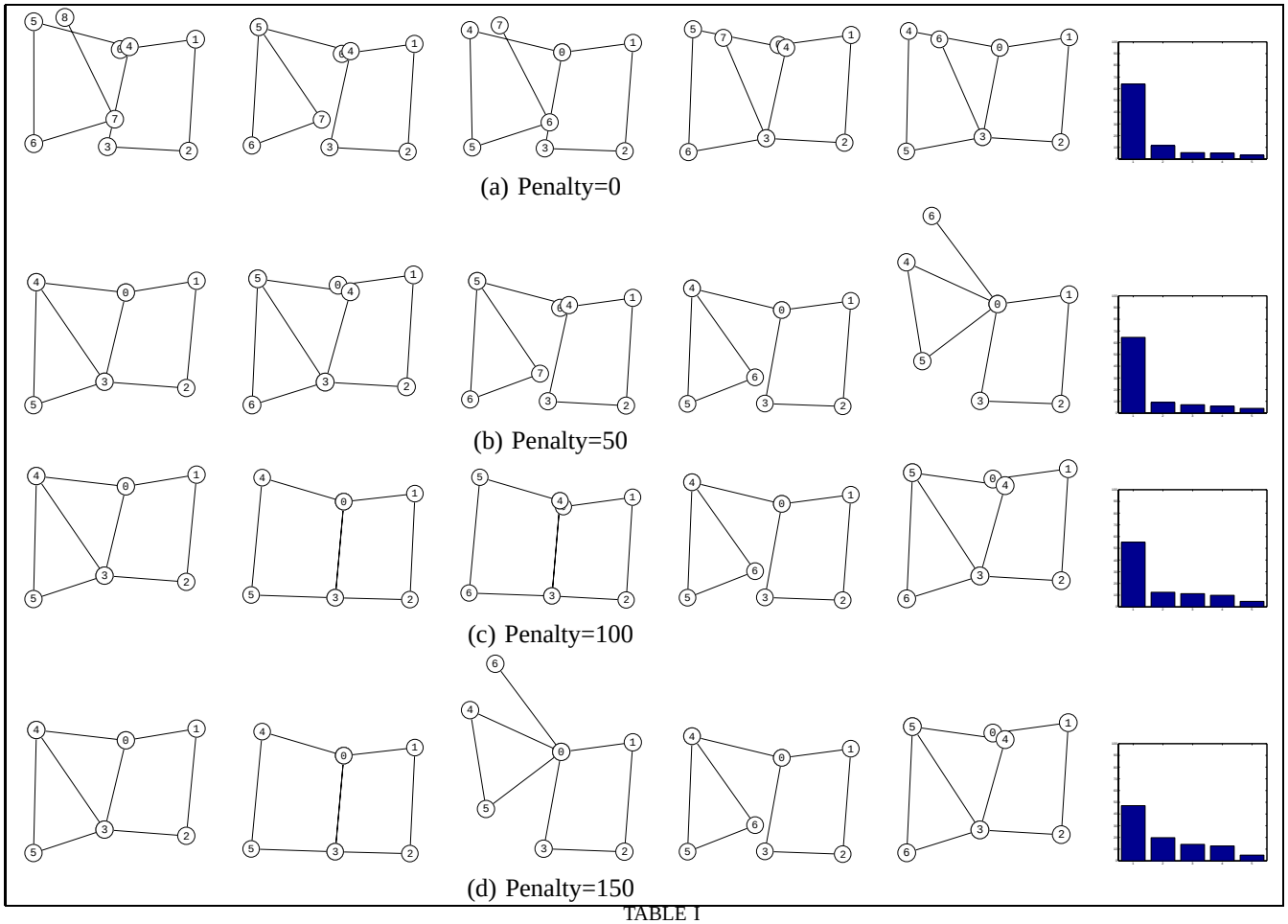
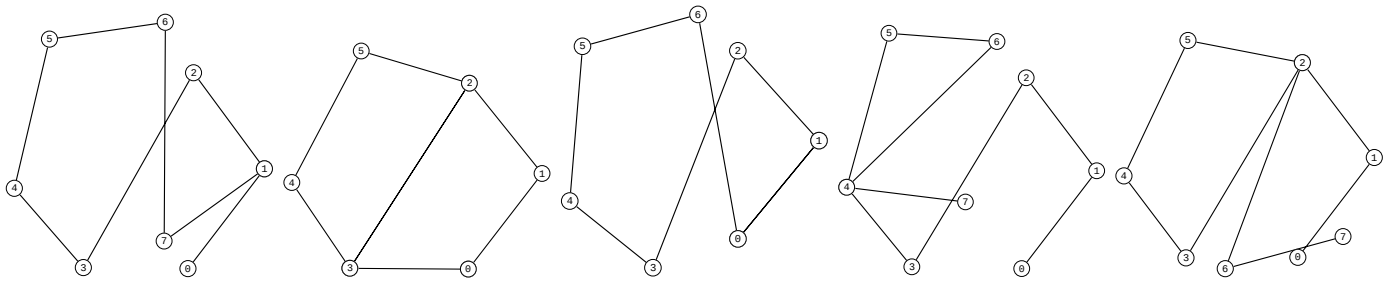


TABLE I

CHANGE IN PROBABILITY MASS WITH MAXIMUM PENALTY OF THE FIVE MOST PROBABLE TOPOLOGIES IN THE HISTOGRAMMED POSTERIOR. THE HISTOGRAM AT THE END OF EACH ROW GIVES THE PROBABILITY VALUES FOR EACH TOPOLOGY IN THE ROW.

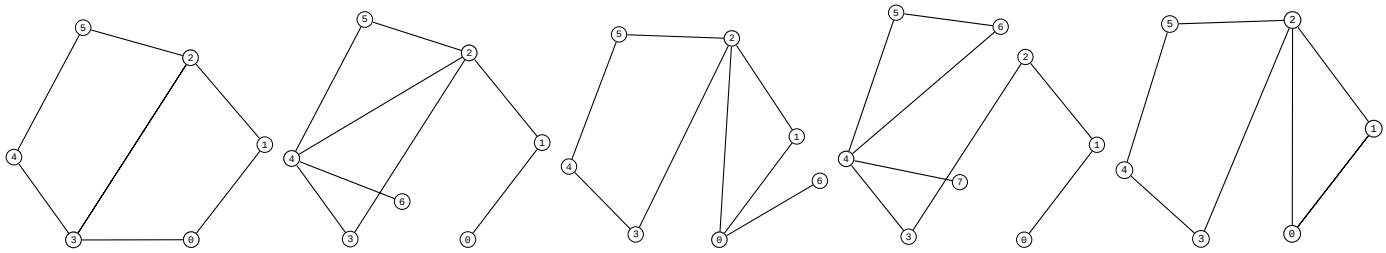


(a) (b) (c) (d) (e)

Fig. 8. The five topologies with highest posterior probability mass for the second experiment using only odometry (a) an incorrect topology receives 91% of the probability mass while the ground truth topology (b) receives 6%, (c), (d) and (e) receive 0.9%, 0.8% and 0.7% respectively.

landmark locations obtained using odometry are shown in Figure 7. As in the first experiment, the five most likely topologies from the target distribution were obtained using only odometry measurements. A penalty radius of 20 meters and a maximum penalty of 100 were used to obtain the topologies, which are shown in Figure 8. As before, the ground truth topology receives only a small probability due to noisy odometry.

We now repeat the experiment, but this time using the appearance measurements, i.e. the Fourier signatures of the panoramic images obtained from the landmark locations, in addition to the odometry. The first five frequencies of the Fourier signatures were used for this purpose. The values of the variance hyper-parameters in the appearance model were set so that the prior over the variance is centered at 500 with a variance of 50. The five most likely topologies in the resulting probability histogram are shown in Figure 9.



(a) (b) (c) (d) (e)

Fig. 9. Topologies with highest posterior probability mass for the second experiment using odometry *and* appearance (a) The ground truth topology receives 94% of the probability mass while (b), (c), (d) and (e) receive 3.2%, 1.2%, 0.3% and 0.3% of the probability mass respectively.

The ground truth topology gets the majority of the probability mass. This experiment illustrates the use of appearance measurements to disambiguate noisy odometry data. Additionally, it demonstrates that the Bayesian model used herein refines the posterior over topologies given more data.

VIII. Conclusions

In the first experiment, even though the environment is small, noisy odometry results in the ground truth topology not receiving the highest probability mass. If a maximum likelihood approach were used in this case, the result would just be an incorrect topology. However, using a Bayesian methodology to compute the posterior over topologies yields the most complete and robust result for the given data. Subsequently, the resulting posterior can be post-processed, if necessary, using an application-specific technique to yield a single topology. On the other hand, computations such as planning and topological localization can be performed using the full posterior without rejecting any possibility. For example, plans can be computed on a multiple topological maps sampled from the posterior distribution; each plan being given a confidence rating proportional to the probability of the sampled topology. In contrast to maximum-likelihood or other truncated approaches, our technique allows such general-purpose, application-specific use of the output.

The second experiment illustrates the power of using a Bayesian approach in the sense that good results are obtained even with noisy data, when a large amount of data is available. Initially, due to odometry drift in the large environment, an incorrect topology gets a large majority of the probability mass. However, the inclusion of appearance measurements, which are themselves noisy, in the inference results in a "better" posterior. Note that in a real application we would not have the ground truth topology and consequently, the notion of a better posterior does not make sense. The experiment merely affirms the fact that use of more data from varied sources improves Bayesian inference and yields more robust results.

IX. Discussion

We presented the novel idea of computing discrete probability densities over the space of all possible topological maps. The Probabilistic Topological Maps are computed using Markov Chain Monte Carlo sampling over set partitions that are used to encode the topologies. PTMs are a systematic solution to the correspondence problem in topological mapping and provide an optimal estimation of the posterior distribution over topologies for the given measurements. We provide a general framework for estimating the posterior over the space of topologies and two specific models for computing measurement likelihood, one that uses odometry and the other using appearance. The odometry likelihood computation uses a simple spatial prior on landmark distribution in the form of a cubic penalty function that disallows proximity among landmarks. The appearance model used in this work deals with Fourier signatures of panoramic images. The model clusters similar appearance measurements as coming from the same spatial location. Experimental results on environments with varied sizes demonstrate the applicability of PTM.

Currently, the PTM algorithm requires four parameters to be chosen by the user. These are the penalty radius and maximum penalty values for the odometry likelihood, and the α and β variance hyper-parameters and the number of frequency components for the appearance likelihood. The penalty values depend on the size and scale of the environment being mapped and need to be empirically determined for each environment. This is also the case for the variance hyper-priors, which encode the variation in appearance values from the same location in the environment. Changes in lighting, uncalibrated cameras and other measurement noise may make this variation large. It is our experience that there is rarely need to use more than the first five frequency components in the appearance model. This is because the higher frequency components mainly contain noise, which we do not seek to model. It is also to be noted that while we use Fourier signatures in this work, any other rotation-invariant dimensionality reduction technique can be used in its stead.

While we only provide likelihood models for odometry and appearance, a simple extension to laser data is also possible. If two lasers are used to gather 360^0 laser scans at the landmark locations, the likelihood of two scans being from the same location can be computed by finding the difference between the scans after an optimal alignment. This likelihood, extended to multiple scan comparison, can be used to sample over partitions.

One advantage of our approach is that an estimate of topology is possible even if only a meager amount of information is available. It is not the purpose of this work to

find the best topological map, but to compute the posterior probability density over topological space as per the Bayesian approach. We have shown this capability in experiments that use only odometry from the robot to create distributions that can either correspond to the odometry or the prior (in this case the spatial penalty function) as parameters are varied. Appearance modeling has largely been used in this work as a disambiguation mechanism for odometry, i.e. by creating more or by decreasing the evidence for the odometry. Of course, more sophisticated appearance models are possible that could be used to completely replace odometry, but that is not the purpose of our appearance model.

A problem with the current setup is the use of a single value for the penalty radius. This can cause poor performance if the distribution of landmarks varies across the environment, for example, if most of the landmarks occur in a closely-spaced group but the remainder are spread wide apart. Finding clusters at different scales is a well-researched problem in machine learning and it is future work to apply those techniques to automate the process of setting the penalty radius.

ACKNOWLEDGMENTS

We would like to thank S. Charles Brubaker for his earlier work on this project, as well as Ruth Conroy-Dalton for many helpful discussions on spatial priors. We are also grateful to Michael Kaess for providing the data used in the experiments.

APPENDIX

To obtain the expression for the appearance likelihood given in (22), consider the integral in (7) which is the probability of a set in the topology taking into account only one frequency component

$$P(S) = \int_{\sigma_{ik}^2, y_{ik}} P(\sigma_{ik}^2) P(y_{ik}|\sigma_{ik}^2) \prod_{j=1}^{|S_i|} P(a_{ijk} \mid y_{ik}, \sigma_{ik}^2)$$

Plugging in the functional forms of the distributions defined in the model (20), we get

$$P(S) = K_i \int_{\sigma_{ik}^2} \left(\sigma_{ik}^2\right)^{-A_i} e^{-\frac{\beta}{\sigma_{ik}^2}} \int_{y_{ik}} e^{-\frac{1}{2\sigma_{ik}^2} B_{ik}}$$

where

$$\begin{aligned} K_i &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\kappa^{\frac{1}{2}}}{(2\pi)^{\frac{|S_i|+1}{2}}} \\ A_i &= \alpha + \frac{|S_i|}{2} + \frac{3}{2} \\ B_{ik} &= \kappa \left(y_{ik} - \mu\right)^2 + \sum_{j=1}^{|S_i|} \left(a_{ijk} - y_{ik}\right)^2 \end{aligned}$$

Performing the inner integration, we get

$$P(S) = K' \int_{\sigma_{ik}^2} \left(\sigma_{ik}^2\right)^{-\gamma_i} e^{-\frac{1}{\sigma_{ik}^2}\left(\beta + \frac{1}{2}\Phi_{ik}\right)} \qquad (23)$$

where

$$\begin{aligned} K' &= \frac{1}{(2\pi)^{\frac{|S_i|}{2}}} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{\kappa}{\kappa + |S_i|}\right)^{\frac{1}{2}} \\ \Phi_{ik} &= \kappa \left(\mu^\star - \mu\right)^2 + \sum_{j=1}^{|S_i|} \left(a_{ijk} - \mu^\star\right)^2 \\ \mu^\star &= \frac{\kappa\mu + \sum_{j=1}^{|S_i|} a_{ijk}}{\kappa + |S_i|} \\ \gamma_i &= \alpha + \frac{|S_i|}{2} + 1 \end{aligned}$$

We now provide here a useful definition of the Gamma function

$$\int_0^\infty e^{-\alpha t} t^\gamma dt = \frac{\Gamma(\gamma+1)}{\alpha^{(\gamma+1)}}$$

using which (23) can be integrated (note that t corresponds to σ_{ik}^{-2}) to yield

$$P(S) = K' \frac{\Gamma(\gamma_i + 1)}{\left\{\beta + \frac{1}{2}\Phi_{ik}\right\}^{(\gamma_i+1)}}$$

whence (22) follows.

REFERENCES

[1] W. Burgard, D. Fox, H. Jans, C. Matenar, and S. Thrun. Sonar-based mapping of large-scale mobile robot environments using EM. In *International Conference on Machine Learning (ICML)*, Bled, Slovenia, 1999.

[2] J.A. Castellanos and J.D. Tardos. *Mobile Robot Localization and Map Building: A Multisensor Fusion Approach*. Kluwer Academic Publishers, Boston, MA, 2000.

[3] H. Choset and K. Nagatani. Topological simultaneous localization and mapping (SLAM): toward exact localization without explicit localization. *IEEE Trans. on Robotics and Automation*, 17(2), April 2001.

[4] F. Dellaert, S.M. Seitz, C.E. Thorpe, and S. Thrun. Feature correspondence: A Markov chain Monte Carlo approach. In *Advances in Neural Information Processing Systems 13 (NIPS 2000)*, 2001.

[5] F. Dellaert, S.M. Seitz, C.E. Thorpe, and S. Thrun. EM, MCMC, and chain flipping for structure from motion with unknown correspondence. *Machine learning*, 50(1-2):45–71, January - February 2003. Special issue on Markov chain Monte Carlo methods.

[6] G. Dissanayake, H. Durrant-Whyte, and T. Bailey. A computationally efficient solution to the simultaneous localisation and map building (slam) problem. Working notes of ICRA'2000 Workshop W4: Mobile Robot Navigation and Mapping, April 2000.

[7] T. Duckett. A genetic algorithm for simultaneous localization and mapping. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 434–439, 2003.

[8] H.F. Durrant-Whyte, S. Majunder, S. Thrun, M. de Battista, and S. Scheding. A Bayesian algorithm for simultaneous localization and map building. Submitted for publication, 2001.

[9] A. Elfes. Occupancy grids: A probabilistic framework for robot perception and navigation. *Journal of Robotics and Automation*, RA-3(3):249–265, June 1987.

[10] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis*. Chapman and Hall, 1995.

[11] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors. *Markov chain Monte Carlo in practice*. Chapman and Hall, 1996.

[12] A. Griewank. On Automatic Differentiation. In M. Iri and K. Tanabe, editors, *Mathematical Programming: Recent Developments and Applications*, pages 83–108. Kluwer Academic Publishers, 1989.

[13] R. Gutierrez-Osuna and R. C. Luo. Lola: Probabilistic navigation for topological maps. *AI Magazine*, 17(1), 1996.

[14] J.-S. Gutmann and K. Konolige. Incremental mapping of large cyclic environments. In *Proc. of the IEEE Intl. Symp. on Computational Intelligence in Robotics and Automation (CIRA)*, 2000.

[15] D. Haehnel, W. Burgard, D. Fox, and S. Thrun. A highly efficient FastSLAM algorithm for generating cyclic maps of large-scale environments from raw laser range measurements. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2003.

[16] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[17] H. Ishiguro, K. C. Ng, R. Capella, and M. M. Trivedi. Omnidirectional image-based modeling: three approaches to approximated plenoptic representations. *Machine Vision and Applications*, 14(2):94–102, 2003.

[18] N. L. Johnson and S. Kotz. *Urn Models and their Applications*. John Wiley and Sons, 1977.

[19] L.P. Kaelbling, A.R. Cassandra, and J.A. Kurien. Acting under uncertainty: Discrete Bayesian models for mobile-robot navigation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1996.

[20] K. Konolige and J.-S. Gutmann. Incremental mapping of large cyclic environments. In *International Symposium on Computational Intelligence in Robotics and Automation (CIRA'99)*, 1999.

[21] D. Kortenkamp and T. Weymouth. Topological mapping for mobile robots using a combination of sonar and vision sensing. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 979–984, 1994.

[22] B. Kuipers and P. Beeson. Bootstrap learning for place recognition. In *AAAI Nat. Conf. on Artificial Intelligence*, 2002.

[23] B.J. Kuipers and Y.-T. Byun. A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. *Journal of Robotics and Autonomous Systems*, 8:47–63, 1991.

[24] J.J. Leonard and H.F. Durrant-Whyte. Simultaneous map building and localization for an autonomous mobile robot. In *IEEE Int. Workshop on Intelligent Robots and Systems*, pages 1442–1447, 1991.

[25] B. Lisien, D. Morales, D. Silver, G. Kantor, I. Rekleitis, and H. Choset. Hierarchical simultaneous localization and mapping. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 448–453, 2003.

[26] F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. Technical report, Department of Computer Science, York University, April 1997.

[27] M. J. Matarić. A distributed model for mobile robot environment-learning and navigation. Master's thesis, MIT, Artificial Intelligence Laboratory, Cambridge, January 1990. Also available as MIT AI Lab Tech Report AITR1228.

[28] E. Menegatti, T. Maeda, and H. Ishiguro. Image-based memory for robot navigation using properties of the omnidirectional images. *Robotics and Autonomous Systems*. To Appear.

[29] E. Menegatti, M. Zoccarato, E. Pagello, and H. Ishiguro. Image-based Monte Carlo localisation with omnidirectional images. *Robotics and Autonomous Systems*. To Appear.

[30] M. Montemerlo and S. Thrun. Simultaneous localization and mapping with unknown data association using FastSLAM. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2003.

[31] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *AAAI Nat. Conf. on Artificial Intelligence*, 2002.

[32] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges. In *Intl. Joint Conf. on Artificial Intelligence (IJCAI)*, 2003.

[33] H.P. Moravec. Sensor fusion in certainty grids for mobile robots. *AI Magazine*, 9:61–74, 1988.

[34] K. Murphy. Bayesian map learning in dynamic environments. In *Advances in Neural Information Processing Systems (NIPS)*, 1999.

[35] J. Neira and J.D. Tardós. Data association in stochastic mapping using the joint compatibility test. *IEEE Trans. on Robotics and Automation*, 17(6):890–897, December 2001.

[36] A. Nijenhuis and H. Wilf. *Combinatorial Algorithms*. Academic Press, 2 edition, 1978.

[37] D. Pierce and B. Kuipers. Map learning with uninterpreted sensors and effectors. *Artificial Intelligence*, 92:169–229, 1997.

[38] E. Remolina and B. Kuipers. Towards a general theory of topological maps. *Artificial Intelligence*, 152(1):47–104, 2004.

[39] H. Shatkay and L. Kaelbling. Learning topological maps with weak local odometric information. In *Proceedings of IJCAI-97*, 1997.

[40] R. Simmons and S. Koenig. Probabilistic robot navigation in partially observable environments. In *Proc. International Joint Conference on Artificial Intelligence*, 1995.

[41] R. Smith, M. Self, and P. Cheeseman. A stochastic map for uncertain spatial relationships. In *Int. Symp on Robotics Research*, 1987.

[42] R. Smith, M. Self, and P. Cheeseman. Estimating uncertain spatial relationships in Robotics. In I. Cox and G. Wilfong, editors, *Autonomous Robot Vehicles*, pages 167–193. Springer-Verlag, 1990.

[43] B. Stewart, J. Ko, D. Fox, and K. Konolige. The revisiting problem in mobile robot map building: A hierarchical Bayesian approach. In *Proceedings of the 19th Annual Conference on Uncertainty in Artificial Intelligence (UAI-03)*, pages 551–558, 2003.

[44] S. Thrun. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence*, 99(1):21–71, 1998.

[45] S. Thrun. A probabilistic online mapping algorithm for teams of mobile robots. *International Journal of Robotics Research*, 20(5):335–363, 2001.

[46] S. Thrun, D. Fox, and W. Burgard. A probabilistic approach to concurrent mapping and localization for mobile robots. *Machine learning*, 31:29–53, 1998.

[47] S. Thrun, S. Gutmann, D. Fox, W. Burgard, and B. Kuipers. Integrating topological and metric maps for mobile robot navigation: A statistical approach. In *AAAI*, pages 989–995, 1998.

[48] L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal distributions. *Journal of the American Statistical Association*, 81:82–86, 1986.

[49] N. Tomatis, I. Nourbakhsh, and R. Siegwart. Hybrid simultaneous localization and map building: Closing the loop with multi-hypotheses tracking. In *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, 2002.

[50] B. Yamauchi and R. Beer. Spatial learning for navigation in dynamic environments. *IEEE Transactions on Systems, Man, and Cybernetics-Part B, Special Issue on Learning Autonomous Robots*, 26:496–505, 1996.

[51] B. Yamauchi and P. Langley. Place recognition in dynamic environments. *Journal of Robotic Systems*, 14(2):107–120, February 1997. Special Issue on Mobile Robots.