Dirichlet Process based Bayesian Partition Models for Robot Topological Mapping

Ananth Ranganathan and Frank Dellaert
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332
{ananth, dellaert}@cc.gatech.edu

Abstract

Robotic mapping involves finding a solution to the correspondence problem. A general purpose solution to this problem is as yet unavailable due to the combinatorial nature of the state space. We present a framework for computing the posterior distribution over the space of topological maps that solves the correspondence problem in the context of topological mapping. Since exact inference in this space is intractable, we present two sampling algorithms that compute sample-based representations of the posterior. Both the algorithms are built on a Bayesian product partition model that is derived from the mixture of Dirichlet processes model. Robot experiments demonstrate the applicability of the algorithms.

1 Introduction

Mapping an unknown and uninstrumented environment is one of the foremost challenges in robotics. In this paper we deal with topological maps [12], which though less popular than their metric counterparts, provide a light-weight, graph-based representation of the environment and hence are useful in many situations. We assume the commonly used definition of a topological map as a graph where the nodes correspond to landmarks in the environment and edges represent the connectivity between landmarks. Possibly the hardest problem in robotic mapping is the correspondence problem, referred to in the case of topological mapping as the perceptual aliasing problem. The aliasing problem in topological mapping involves matching sensor measurements to physical locations from which they were obtained. Solutions to the this problem have to deal with a state space that grows combinatorially with measurement size.

In this paper, we present a framework for computing the posterior distribution over the space of topological maps. We use the fact that a topology can be viewed as a set partition over the set of measurements; a set in the partition being the measurements corresponding to a physical landmark in the environment. Hence, the space of possible topologies is isomorphic to the space of set partitions. A Product Partition Model (PPM) derived from the Mixture of Dirichlet Processes (MDP) model is used to perform inference over the space of topologies. We present two algorithms to infer the posterior distribution: a Markov Chain Monte Carlo algorithm with split-merge proposals, and a Rao-Blackwellized Sequential Importance Sampling algorithm.

The key contribution of this paper is the idea of defining a probability distribution over topologies and the use of a product partition model framework to infer it. The intuitive reason for computing the posterior is to solve the aliasing 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, whereby each topology is assigned a probability, it is possible to locate the more probable topologies without committing to a specific correspondence at any point in time, thus providing the most general solution to the aliasing problem. Even in pathological environments, where almost all current algorithms fail, our technique provides a quantification of uncertainty by pegging a probability of correctness to each topology. The solution of the perceptual aliasing problem by sample-based estimation of the posterior over topological maps is completely novel to the best of our knowledge.

As an additional contribution, we provide a summary of the use of the MDP model as a PPM, which is previously lacking in literature. The role of the Dirichlet Process as a prior over all possible data clusterings (or set partitions) is clearly illustrated. Our experiments use models of appearance and odometry to obtain discrete, histogram representations of the posterior distribution over the space of topologies.

Previous work in the area of inference over topological maps mainly consists of the use of hidden and partially observable markov models to solve the correspondence problem, for example [13, 15]. A Dirichlet prior on maps has recently been used by Stewart et. al. [14] to model 'views' that the robot sees when using different maps. The use of the MDP model as a PPM has been enunciated most recently in [11].

2 Topological Mapping using PPMs

The aim of this work is to compute the posterior over the space of topologies. To this end, we assume the availability of a set of appearance measurements $Y = \{y_i | 1 \leq i \leq n\}$. Using Bayes Law, the posterior over topology Θ given measurements Y can be written as

$$p(\Theta|Y) \propto p(Y|\Theta)p(\Theta) \tag{1}$$

where $p(\Theta)$ is a prior over the space of topologies. Due to the equivalence between a topology and a set partition, Θ also defines a set partition given as $\Theta = \{S_1, S_2, \ldots, S_m\}$ s.t. $S_i \cap S_j = \phi, \forall 1 \leq i, j \leq m$, $i \neq j$ and $S = \bigcup_{i=1}^{m} S_i$ on the set $S = \{1, 2, \cdots, n\}$. Alternately, Θ defines a partition where every set indexes the measurements in Y that correspond to the same landmark.

Appearance measurements in a particular set of the partition are independent of measurements in other sets, since they only depend only on the corresponding physical landmark. If in addition, we assume that the prior on Θ can be factored into the prior probabilities of the individual sets, we obtain a further factorization of (1) as

$$p(\Theta|Y) \propto \prod_{i=1}^{m} p(y_{S_i})p(S_i) \tag{2}$$

where $p(y_{S_i})$ is the set likelihood of measurements indexed by the set S_i.

PPMs, first introduced by Hartigan [5], have exactly the same form as (2) due to their assumption of independence between sets in the partition. Hence, the posterior over topologies is, in fact, a product partition model. Another important characteristic of the PPM is that measurements in each set of the partition are exchangeable. This can be seen from (2) which does not specify an ordering on the y_{S_i} for calculating the set likelihood $p(y_{S_i})$.

3 Mixture of Dirichlet Processes as a Product Partition Model

The previous section demonstrated that the posterior over topologies can be viewed as a PPM. However, specific forms for the set likelihood and prior distributions in (2) remain to be specified. In this section, we show that the mixture of Dirichlet processes (MDP) model can be used as a PPM that provides intuitive forms for these distributions, and hence, is well suited for topological mapping.

In order to obtain a PPM from a MDP, we first consider the standard MDP model [1]

$$\begin{aligned} y_j|\lambda_j &\sim p(\cdot|\lambda_j) \\ \lambda_j &\sim G \\ G &\sim \mathcal{DP}(\alpha G_0) \end{aligned} \tag{3}$$

where $\mathcal{DP}(\alpha G_0)$ is the Dirichlet Process centered around the distribution G_0 with precision parameter α, and $p(\cdot|\lambda_j)$ is a family of distributions indexed by λ_j. The MDP model explains the measurements are arising from a mixture of distributions of the same form $p(\cdot|\cdot)$. The parameters of each mixture component are obtained as samples from a random probability distribution G, which, in turn, is a sample from the Dirichlet Process. Blackwell and MacQueen [2] showed that the distribution G in (3) could be integrated out to yield a Polya Urn scheme for the λ_j, given as

$$\begin{aligned} y_j|\lambda_j &\sim p(\cdot|\lambda_j) \\ \lambda_j|\lambda_{1:j-1} &\sim \frac{\alpha G_0 + \sum_{k=1}^{j-1} \delta(\lambda_k)}{\alpha + j - 1} \end{aligned} \tag{4}$$

where $\delta(\lambda_k)$ is a point probability mass at λ_k.

The Polya Urn scheme provides a distribution over the possible correspondences for each new measurement. The scheme gives the probability that the measurement arises from one of the previously observed parameter values λ_k s.t. $k \in [1, j-1]$, or from a new parameter that is drawn from the distribution G_0. In the context of topological mapping, this is equivalent to the probability of the measurement corresponding to one of the previously observed landmarks or a new landmark, which is exactly the distribution over correspondences for this measurement. This crucial observation provides the intuition linking topological mapping and the MDP model.

To rewrite the MDP model as a PPM, we note that the vector $\Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ can be re-parametrized as $\{\Theta, \Phi\}$, where Θ is a set partition as defined above, and $\Phi = \{\phi_1, \phi_2, \cdots, \phi_m\}$ is a vector of set parameters, with every ϕ_i being paired with the corresponding S_i in Θ. The marginal distribution of elements in S is again given by a Polya Urn formulation similar to (4). The prior over partitions $p(\Theta)$ can be obtained by calculating the joint distribution over the elements in S

$$p(\Theta) = \frac{\alpha^m}{\prod_{j=1}^{n} (\alpha + j - 1)} \prod_{i=1}^{m} (|S_i| - 1)! \tag{5}$$

and the new parametrization allows us to rewrite the model in (3) as

$$\begin{aligned} y_j|\Theta, \Phi &\sim p\left(\cdot|\phi_{S\{j\}}\right) \\ \Theta &\sim p(\Theta) \quad \text{given in (5)} \\ \phi_{S\{j\}} &\sim G_0 \end{aligned} \tag{6}$$

where $S\{j\}$ represents the set in Θ that contains j, and $\phi_{S\{j\}}$ are the parameters for this set.

The likelihood of the data Y given Θ can be obtained from (6) by marginalizing over the ϕ_i, which gives us the expression for the likelihood as

$$p(Y|\Theta) = \prod_{i=1}^{m} \int_{\phi_i} p(y_{S_i}|\phi_i) G_0(\phi_i) \tag{7}$$

where y_{S_i} is the set of observed data corresponding to the set S_i.

On comparing the definition of the PPM in (2) with the model given by (5) and (7), it is clear that the reparametrization of the MDP model results in a PPM. A more rigorous treatment of this topic is available in [3]. Given the above framework, we use (5) as the prior over topologies and (7) to evaluate the appearance likelihood in (1).

4 Incorporating Odometry

The above discussion neglects to mention odometry, which is an essential part of any robotic mapping scheme. We rectify the situation in this section.

In addition to the set of appearance measurements Y , we now also assume the availability of a set of odometry measurements $O = \{o_i | 1 \leq i \leq n-1\}$. This set has only $n-1$ measurements as the robot is assumed to start from the origin. Further, we demonstrate in section 7 that our appearance measurements are rotationally invariant, so that we can assume conditional independence between the odometry and appearance data given the topology. Using this conditional independence, we can add O to (1) to obtain the posterior over topologies as

$$p(\Theta|O, Y) \propto p(O|\Theta)p(Y|\Theta)p(\Theta) \tag{8}$$

However, odometry cannot be treated the same way as appearance measurements. This is due to the fact that odometry measurements are clearly not exchangeable since they give the distance and direction traveled by the robot between landmarks and hence, have to be considered in sequence. Thus, a factorization that avails the independence between sets in the partition is clearly not possible.

We sidestep the problem of finding independence conditions among the odometry measurements by simply calculating the joint odometry likelihood without any factorization. Moreover, it is not possible to compute the joint likelihood without knowledge of the landmark locations. Since we are not interested in the landmark locations, we marginalize over them

$$p(O|\Theta) = \int_X p(O|X, \Theta)p(X|\Theta) \tag{9}$$

where $X = \{x_1, \cdots, x_n\}$ is the vector of landmark locations with x_1 being fixed as the origin, and $p(X|\Theta)$ is a prior over the distribution of landmarks in the environment. The physical meaning of (9) is that we consider all possible vectors of landmark locations that could account for the odometry measurements.

The posterior in (8) is analytically intractable as the state space grows hyper-exponentially with the number of measurements. Consequently, we compute sample-based representations of the posterior using sampling algorithms described in the following sections.

5 Inferring the Posterior over Topological Maps using MCMC

This section describes the design of an MCMC sampler based on the Metropolis-Hastings (MH) algorithm for inferring the posterior in (8). The sampler incorporates a split-merge proposal distribution, which chooses one of the split and merge steps with equal probability, to compute the posterior over the space of topologies. The acceptance ratio calculation closely follows [7] and involves Rao-Blackwellization as the parameters are integrated away.

The split step works by randomly picking a set from the topology and generating an arbitrary split of this set. If the number of non-singleton sets in the topology is N_S and the cardinality of the set chosen to be split at random is R , then the probability of the split is given by $\left(N_S \left\{ {|R| \atop 2} \right\}\right)^{-1}$, where $\left\{ {|R| \atop 2} \right\}$ is the Stirling number of the second kind that gives the number of possible splits of a set of cardinality R into two sets. Since sets in the partition correspond to landmarks in the topology, the split step results in a topology with one more landmark node than before.

For the *merge step*, we similarly pick two sets at random and propose to merge them, the probability of the merge being $\binom{N}{2}^{-1}$, where N is the number of sets in T , and $\binom{N}{2}$ is the binomial coefficient. This step results in a topology with one less landmark than before. In either case, if a chosen step is not possible, the current topology is re-proposed.

The MH acceptance probability, considering the current topology to be Θ and the new topology to be $\Theta^\star$, is given by

$$a(\Theta^\star, \Theta) = \min\left(1, \frac{q(\Theta|\Theta^\star)}{q(\Theta^\star|\Theta)} \frac{P(\Theta^\star)}{P(\Theta)} \frac{P(O|\Theta^\star)P(Y|\Theta^\star)}{P(O|\Theta)P(Y|\Theta)}\right) \tag{10}$$

where $q(\cdot|\cdot)$ is the proposal probability calculated above and use has been made of (8) to expand the posterior distribution. Note that $\Theta^\star$ is one of Θ^{split} and Θ^{merge}.

The prior ratio $\frac{p(\Theta^\star)}{p(\Theta)}$ in (10) is calculated using (5). Let S_i and S_j be the result of splitting a random set S in Θ during a split step and conversely, let S be the result of merging S_i and S_j in a merge step. Then, cancelling common terms in the prior ratio, we get

$$\frac{p(\Theta^{split})}{p(\Theta)} = \left(\frac{p(\Theta^{merge})}{p(\Theta)}\right)^{-1} = \alpha\frac{(|S_i|-1)!\,(|S_j|-1)!}{(|S|-1)!}$$

The ratio of the appearance likelihood can be calculated using (7), where again terms corresponding to sets common to Θ and $\Theta^\star$ cancel, to yield

$$\frac{p(Y|\Theta^{split})}{p(Y|\Theta)} = \left(\frac{p(Y|\Theta^{merge})}{p(Y|\Theta)}\right)^{-1} = \frac{\int_{\phi_i} p(y_{S_i}|\phi_i)p(\phi_i) \times \int_{\phi_j} p(y_{S_j}|\phi_j)p(\phi_j)}{\int_\phi p(y_S|\phi)p(\phi)} \tag{11}$$

The ratio of the odometry likelihood is evaluated using (9), and no simplification is possible in this case.

6 Sequential Importance Sampling for Inferring the Posterior over Topological Maps

While the MCMC algorithm presented above works well in most cases, it has a few problems. Local maxima in the posterior may cause the chain to mix slowly resulting in poor performance. Though data-driven proposals, such as the one given in [7], can be incorporated to overcome this problem, this slows down the algorithm considerably. In this section, we present a Sequential Importance Sampling (SIS) algorithm that uses Rao-Blackwellization and often outperforms the MCMC algorithm in such situations.

We use a slight modification of the Rao-Blackwellized SIS algorithm given in [9] to compute a sample-based version of the posterior. While ordinary SIS algorithms for MDP models incorporate location parameters in the sampling, Rao-Blackwellization improves performance by marginalizing out these variables. The weight calculation involves sequential imputation as described in [8].

To derive the expressions for the prediction step and importance sampling weights, we start by rewriting (7) in the following form

$$p(Y|\Theta) \quad = \quad \prod_{i=1}^{m}\prod_{j=1}^{|S_i|}\int_{\phi_i} p(y_{i,j}|\phi_i)p\left(\phi_i|y_{i,1}, y_{i,2}, \cdots, y_{i,j-1}\right) \tag{12}$$

where $y_{i,j}$ is the jth measurement in set S_i, and the expression $p\left(\phi_i|.\right)$ is the posterior on ϕ_i taking into account all previous measurements belonging to the set S_i. For $j = 1$, this distribution is the same as G_0. (12) , in turn, can be used to obtain the likelihood of an individual measurement, given the set to which it belongs

$$p(y_i|s_i = j, y_{1:i-1}, s_{1:i-1}) \quad = \quad \int_\phi p(y_i|\phi)p\left(\phi|y_{1:i-1,s_i}\right) \tag{13}$$

where s_i is an indicator variable that gives the set containing the ith measurement y_i, $y_{1:i-1,s_i}$ is the set of measurements $\{y_k|k < i, s_k = s_i\}$ that have the same indicator as s_i, and $y_{1:i-1}$ denotes all the measurements $\{y_k|1 \leq k \leq i-1\}$.

The prediction/proposal step for the SIS filter is obtained by combining (13) with the conditional distribution on s_i, given by the Polya Urn scheme

$$p(s_i = j|s_{1:i-1}) = \frac{\alpha I(n_j = 0) + n_j}{\alpha + i - 1} \tag{14}$$

where $n_j = \sum_{k=1}^{i-1} \delta(s_k = j)$ and I is the indicator function returning unity if its argument is true and zero otherwise. Assuming $D_{i-1} = \{y_{1:i-1}, s_{1:i-1}\}$, the prediction step is obtained as

$$p(s_i = j|D_{i-1}, y_i) \propto p(y_i|s_i = j, D_{i-1})p(s_i = j|s_{1:i-1}) \tag{15}$$

The predictive appearance likelihood, used in the calculation of weights, can be obtained from (13) and (14)

$$p(y_i|y_{1:i-1}, s_{1:i-1}) \propto \alpha \int_\phi p(y_i|\phi)G_0(\phi) + \sum_{j=1}^{K} n_j \int_\phi p(y_i|\phi)p\left(\phi|y_{1:i-1,j}\right) \tag{16}$$

where K is the number of distinct values contained in $s_{1:i-1}$(or alternately, the number of sets in the first $i-1$ observations).

The expression for the weights in the SIS algorithm is derived using sequential imputation, the details of which are given in Appendix A. The algorithm consists of the following steps

1. Repeat (a) and (b) to obtain R samples, then normalize the weights
   (a) For $i = 1, \ldots, n$ do
      i. Generate s_i from the multinomial distribution (15)
      ii. Calculate $w_i = p(y_i|s_{1:i-1}, y_{1:i-1})$ from (16)
   (b) Calculate the importance sampling weight for the sample $\{s_1, s_2, \ldots, s_n\}$ as $w_r = p(o_{1:n-1}|s_{1:n}) \prod_{i=1}^{n} w_i$

7 Experiments

The previous sections formulated a general purpose framework for topological mapping using appearance and odometry. In this section, we describe the specific models used by us and the experiments performed using these models.

We use Fourier signatures [6] of images, obtained from an eight camera rig mounted on the robot, as appearance measurements. A Fourier signature is a row-wise 1D Fourier transform of an image, widely used as a low-dimensional representation of omni-directional images. A Gaussian model with unknown mean and variance is assumed for these measurements. In this model, the set location parameters are given as $\phi = \{\mu, \sigma^2\}$. We assume a conjugate prior over μ and σ^2, so that the base measure of the Dirichlet process G_0 is a Normal-Inv-Gamma distribution. The specific MDP model used, corresponding to (6) is

$$\begin{aligned} y_i|\Theta, \Phi &\sim \mathcal{N}\left(\mu_{S\{i\}}, \sigma^2_{S\{i\}}\right) \\ \Theta &\sim p(\Theta) \quad \text{given in (5)} \\ \mu_{S\{i\}}|\sigma^2_{S\{i\}} &\sim \mathcal{N}\left(\psi, \frac{\sigma^2_{S\{i\}}}{\kappa}\right) \\ \sigma^2_{S\{i\}} &\sim IG(a, b) \end{aligned} \tag{17}$$

where $IG(.)$ is the inverse gamma distribution. The hyperprior ψ is learned from data while κ, a and b are given subjective values. The a and b parameters encode the variation in the appearance measurements obtained from the same location. Use of the conjugate prior allows the evaluation of the acceptance ratio in (10) and weight computation in (16)

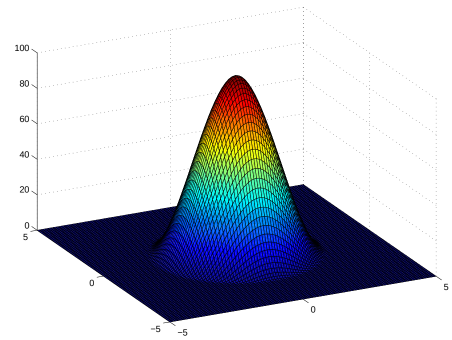


Figure 1: The prior on landmark locations used in the odometry model.

to be performed analytically. In the interests of continuity, the details of these calculations are deferred to Appendix B.

For the odometry model, we assume that landmark locations have the 2D form $\{x, y, \theta\}$, as is common in the robotics literature. A simple prior on landmarks, required in (9), is assumed, and encodes the assumption that landmarks do not exist close together in the environment. A topology Θ placing two distinct landmarks x_i and x_j within a distance d of each other receives a "penalty", given by the penalty function $L(x_i, x_j) = f(d)I(d < D)$, where d is the euclidean distance between x_i and x_j, I is the indicator function which is 0 if $d \geq D$, D is a threshold distance called the "penalty radius", and we define $f(d)$ to be a cubic function with value 0 at $d = D$ and maximum penalty value M at $d = 0$ as shown in Figure 1. The total probability of landmark locations X given topology Θ , $p(X|\Theta)$, can then be calculated as

$$p(X|\Theta) = \prod_{\substack{1 \leq i < j \leq n \\ x_j \notin S\{x_i\}}} e^{-L(x_i, x_j)} \tag{18}$$

where $S\{x_i\}$ denotes the set containing x_i.

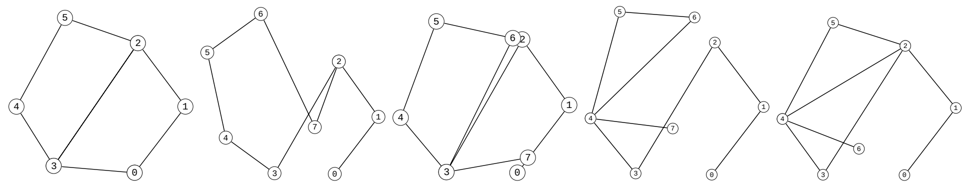
The integration in (9) cannot be performed analytically as the form of $p(O|X, \Theta)$ is unknown. Instead, we use non-linear minimization to find a mode in the integrand in (9) and subsequently, employ Laplace's approximation to calculate the integral. In practice, the minimization is performed using the Levenberg-Marquardt algorithm in an Automatic Differentiation framework [4]. The objective function to be minimized, which encodes the error between the measured odometry and the topology Θ as well as the error due to the landmark prior, is defined to be

$$\psi(X) = \left(\frac{X - X_O}{\sigma_O}\right)^2 + \sum_{S \in T} \sum_{i,j \in S} \left(\frac{x_i - x_j}{\sigma_\Theta}\right)^2 - \log p(X|\Theta) \tag{19}$$

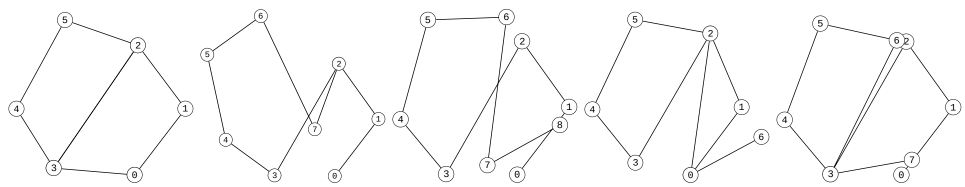
where X_O are the landmark locations obtained from the odometry, σ_O is the variance of the measurement error in the odometry, σ_Θ encodes the distortion in the topology that can be tolerated, and $p(X|\Theta)$ is as defined in (18).

8 Results

Robot experiments were performed using an ATRV-Mini mounted with an eight-camera rig. The images from the rig were automatically aligned to obtain an omni-directional view

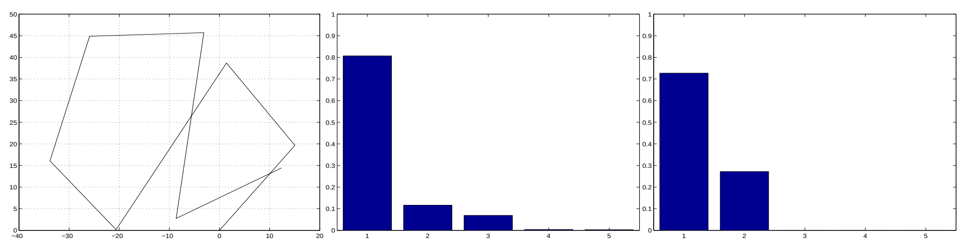


(a)



(b)

Table 1: The five most probable topologies in the posterior (obtained as a histogram) (a) using $\alpha = 1$ (b) $\alpha = 0.1$.



(a) (b) (c)

Table 2: (a)Landmark locations plotted using odometry for the experimental environment, (b) Histogram giving probability masses for the topologies in Figure 1(a): the last two topologies have probability 3% and 2%. (c) The corresponding histogram for the topologies in Figure 1(b): the last three topologies all have negligible probabilities less than 0.1%.

at each location. Experiments were conducted in an indoor office environment with nine landmarks. Raw odometry from the robot, giving the scale of the environment, is shown in Figure 2(a). Histogram representations of the posterior distribution were constructed as output. Figures 1(a) and 1(b) depict the five most probable topologies in the histogram obtained from the MCMC and SIS algorithms on this data for two values of the α parameter. The histogram values themselves are provided in Figures 2(b) and 2(c). Note that both the algorithms produce the same topologies since they sample from the same posterior.

The MCMC and SIS algorithms take about 15 and 9 seconds respectively, to generate 15000 samples on a 2 GHz machine. For the MCMC algorithm, a burn-in period of 5000 samples was used. In all the experiments, the hyper-parameters κ, a and b were given general values of 1, 2 and 3 respectively. The maximum penalty value used was 200, while the penalty radius was set to 25 meters.

The algorithms are robust to the values of the hyper-parameters. This was tested by changing the hyper-parameter values by a factor of 10, which results in no drastic changes in the probability histogram. For smaller values of α, the algorithms tend to converge around one particular topology and the histograms lose diversity. This is not surprising as for smaller values of α , the inferred mixture model from the MDP tends to become more peaked, so that finding all the modes of the distribution becomes difficult.

9 Discussion

We demonstrated the use of the MDP based PPM for solving the perceptual aliasing problem in topological mapping in a robust, Bayesian manner. The framework we presented is completely novel in the domain of topological mapping. The result of our algorithm is the posterior distribution over topologies, which represents the most general solution possible to the topological mapping problem. We derived two sampling algorithms that compute sample-based representations (in our case a histogram) of the posterior. We also showed how to incorporate non-exchangeable odometry measurements into this framework in a natural manner. Experimental results validate our approach and demonstrate the robustness of our algorithm with respect to parameter values.

Our algorithm is a major step towards acknowledging the idea that the ideal mapping algorithm, capable of producing an accurate map in any environment using just the available measurements, may not exist. Instead, the mapping algorithm should be able to reason about and flag any uncertainty it might have about the maps it is generating. This is precisely what our technique accomplish. The posterior over maps produced by our method could also be used as the basis to create a posterior over all possible metric maps using the approach given in [10]. In addition, it might be useful to incorporate it into probabilistic planning problems, where the probability of a map could be reflected in the probability of success of the plan.

In the future, we will make odometry calculations incremental by incorporating a motion model. A systematic technique for finding optimal values for the penalty parameters is also to be explored.

Appendix A

In this appendix, we explain the mechanism behind the SIS algorithm and derive an expression for the weights used therein. We start by noting that the posterior over topologies is equivalent to the joint posterior over the indicator variables $p(s_{1:n}|o_{1:n-1}, y_{1:n})$. This is true since the indicator variables jointly define a set partition over the measurements. hence, our aim is to compute the posterior $p(s_{1:n}|o_{1:n-1}, y_{1:n})$. First, we apply Bayes law and utilize the independence between the odometry and appearance measurements to write the posterior as

$$\begin{aligned} p(s_{1:n}|o_{1:n-1}, y_{1:n}) &= \frac{p(o_{1:n-1}|s_{1:n})p(y_{1:n}|s_{1:n})p(s_{1:n})}{p(o_{1:n-1})p(y_{1:n})} \\ &= \frac{p(o_{1:n-1}|s_{1:n})p(s_{1:n}|y_{1:n})}{p(o_{1:n-1})} \end{aligned} \tag{20}$$

The odometry likelihood can be calculated from (9). Hence, we concentrate hereafter on the posterior given only the appearance measurements $p(s_{1:n}|y_{1:n})$. Since we cannot obtain this posterior analytically, we perform sequential imputation using importance sampling to achieve this.

The importance sampling proposal distribution is defined as

$$p^\star(s_{1:n}|y_{1:n}) = p(s_1|y_1)\prod_{t=2}^{n} p(s_t|y_{1:t}, s_{1:t-1}) \tag{21}$$

We can sample from this proposal distribution since the marginal distributions over the indicator variables in the right side of the equation are given by (15), which in turn, is a multinomial distribution. This procedure is called sequential imputation [8] since the marginals on the indicator variables in (21) are sampled sequentially.

To compute the importance weights, we rewrite (21) as

$$p^\star(s_{1:n}|y_{1:n}) \quad = \quad p(s_1|y_1)\prod_{t=2}^{n}\frac{p(y_{1:t},s_{1:t})}{p(y_{1:t},s_{1:t-1})} \tag{22}$$

The importance weight is given as

$$\begin{aligned} w &= \frac{p(s_{1:n}|y_{1:n})}{p^\star(s_{1:n}|y_{1:n})} \\ &= \frac{p(s_{1:n},y_{1:n})}{p(y_{1:n})}\times\frac{p(y_1)}{p(s_1,y_1)}\times\prod_{t=2}^{n}\frac{p(y_{1:t},s_{1:t-1})}{p(y_{1:t},s_{1:t})} \end{aligned}$$

where use has been made of (22). We continue with some more manipulations

$$\begin{aligned} w &= \frac{p(s_{1:n},y_{1:n})}{p(y_{1:n})}\times\frac{p(y_1)}{p(s_{1:n},y_{1:n})}\times\prod_{t=2}^{n}\frac{p(y_{1:t},s_{1:t-1})}{p(y_{1:t-1},s_{1:t-1})} \\ &\propto p(y_1)\times\prod_{t=2}^{n}p(y_t|s_{1:t-1},y_{1:t-1}) \end{aligned} \tag{23}$$

Combining the above expression for the importance weight with (20), we get the sample-based expression for the posterior as

$$p(s_{1:n}|o_{1:n-1},y_{1:n}) \quad = \quad \frac{\sum_{i=1}^{m} w_i p(o_{1:n-1}|s_{1:n}^{(i)})}{\sum_{i=1}^{m} w_i}$$

where w_i is the weight corresponding to the ith sample $s_{1:n}^{(i)}$ obtained using sequential imputation.

Appendix B

In this appendix, we give the calculations used to obtain the acceptance ratio in the MCMC algorithm and the weights in the SIS algorithm for the appearance models used in our experiments. We only deal with the portion of the calculations pertaining to the appearance measurements. The odometry likelihood cannot be obtained analytically, and the method to compute it is explained in section 7.

We start with the observation that both the acceptance ratio and the weight calculation depend only on the computation of the likelihood of the ith appearance measurement, given that this measurement belongs to a particular set in the partition. In other words, we need to compute $p(y_i|s_i = j, y_{1:i-1,j})$, where $y_{1:i-1,j}$ refers to all the previous measurements in the jth set.

Marginalizing over the parameters for this set, we obtain

$$p(y_i|s_i = j, y_{1:i-1,j}) \quad = \quad \int_\phi p(y_i|\phi_j)p(\phi_j|y_{1:i-1,j}) \tag{24}$$

If we take our appearance measurement model to be (17), the set parameters are the mean and variance of the measurements in the set, $\phi_j = \{\mu_j, \sigma_j^2\}$. The posterior over these parameters is a Normal-Inv-Gamma distribution due to conjugacy

$$p(\phi_j|y_{1:i-1,j}) \quad = \quad \mathcal{N}(\mu_j;\mu,\frac{\sigma^2}{\tau})\mathrm{IG}(a',b') \tag{25}$$

where

$$\begin{aligned}\tau &= \kappa + i - 1 \\ \mu &= \frac{\kappa\psi + \sum_{k=1}^{i-1} y_{k,j}}{\tau} \\ a' &= a + \frac{i-1}{2} \\ b' &= b + \frac{1}{2}\sum_{k=1}^{i-1}(y_k - \mu)^2 + \kappa(\mu - \psi)^2\end{aligned}$$

Using (25) in (24), we get

$$p(y_i|s_i = j, y_{1:i-1,j}) = C\int_{\mu_j,\sigma_j^2} \left(\sigma_j^2\right)^{-a'-2} \exp\left\{-\frac{b'}{\sigma_j^2} - \frac{K}{\sigma_j^2}\right\}$$

where

$$\begin{aligned}C &= \frac{\sqrt{\tau}b'^{a'}}{2\pi\Gamma(a')} \\ K &= \frac{1}{2}\left\{(y_i - \mu_j)^2 + \tau(\mu_j - \mu)^2\right\}\end{aligned}$$

Performing the integration on μ_j, we get

$$p(y_i|s_i = j, y_{1:i-1,j}) = \frac{b'^{a'}}{\Gamma(a')}\sqrt{\frac{\tau}{2\pi(1+\tau)}}\int_{\sigma_j^2} \left(\sigma_j^2\right)^{-a'-\frac{3}{2}} \exp\left(-\frac{\epsilon}{\sigma_j^2}\right) \tag{26}$$

where

$$\begin{aligned}\epsilon &= b' + \frac{1}{2}\left\{(y_i - \mu_j^\star)^2 + \tau(\mu_j - \mu^\star)^2\right\} \\ \mu_j^\star &= \frac{y_i + \mu\tau}{1+\tau} = \frac{y_i + \kappa\psi + \sum_{k=1}^{i-1} y_{k,j}}{\kappa + i}\end{aligned}$$

We 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 we integrate out σ_j^2 from (26) (note that t corresponds to σ_j^{-2}) to yield

$$p(y_i|s_i = j, y_{1:i-1,j}) = \frac{b'^{a'}}{\Gamma(a')}\sqrt{\frac{\tau}{2\pi(1+\tau)}}\frac{\Gamma(\gamma+1)}{\epsilon^{\gamma+1}} \tag{27}$$

where

$$\gamma = a' + \frac{3}{2} = \frac{a+i}{2} + 1$$

We can now use (27) in (15) and (16) to calculate the weights for the samples in the SIS algorithm. Similarly, this can also be used to compute the Metropolis-Hastings acceptance ratio given by (11) since

$$\int_\phi p(y_S|\phi)p(\phi) = \int_\phi p(y_m|\phi)p(\phi|y_{1:m-1,S})$$

where it is assumed that set S has m measurements and as before, y_S is the set of all measurements in S.

References

[1] C. E. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics*, 2:1152–1174, 1974.

[2] D. Blackwell and J.B. MacQueen. Ferguson distributions via polya urn schemes. *Annals of Statistics*, 1:353–355, 1973.

[3] D. B. Dahl. Modal clustering in a univariate class of product partition models. Technical Report 1085, Department of Statistics, University of Wisconsin, 2003.

[4] 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.

[5] J. A. Hartigan. Partition models. *Communications in Statistics, Part A - Theory and Methods*, 19:2745–2756, 1990.

[6] H. Ishiguro and S. Tsuji. Image-based memory of environment. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 1996.

[7] S. Jain and R. Neal. A split-merge Markov chain Monte Carlo procedure for the dirichlet process mixture model. *Journal of Computational and Graphical Statistics*, 13(1):158–182, March 2004.

[8] A. Kong, J. S. Liu, and W. H. Wong. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89(425):278–288, 1994.

[9] S. N. MacEachern, M. Clyde, and J.S.Liu. Sequential importance sampling for nonparametric Bayes models:the next generation. *Canadian Journal of Statistics*, 27:251–267, 1999.

[10] J. Modayil, P. Beeson, and B. Kuipers. Using the topological skeleton for scalable global metrical map-building. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2004.

[11] Fernando A. Quintana and Pilar L. Iglesias. Bayesian clustering and product partition models. *J Royal Statistical Soc B*, 65(2):557–557, 2003.

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

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

[14] 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.

[15] 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.