The International Journal of
Robotics Research
00(000) 1–17

DOI: 10.1177/0278364910393287
ijr.sagepub.com

Online Probabilistic Topological Mapping

Ananth Ranganathan[1] and Frank Dellaert[2]

Abstract

We present a novel algorithm for topological mapping, which is the problem of finding the graph structure of an environment from a sequence of measurements. Our algorithm, called Online Probabilistic Topological Mapping (OPTM), systematically addresses the problem by constructing the posterior on the space of all possible topologies given measurements. With each successive measurement, the posterior is updated incrementally using a Rao–Blackwellized particle filter. We present efficient sampling mechanisms using data-driven proposals and prior distributions on topologies that further enable OPTM's operation in an online manner. OPTM can incorporate various sensors seamlessly, as is demonstrated by our use of appearance, laser, and odometry measurements. OPTM is the first topological mapping algorithm that is theoretically accurate, systematic, sensor independent, and online, and thus advances the state of the art significantly. We evaluate the algorithm on a robot in diverse environments.

Keywords

1. Introduction

Topological mapping is the problem of computing a graphical structure of the environment from sensor measurements. In general, the nodes of the topological graph represent landmarks or significant places while the edges denote connectivity between the landmarks. This is a more abstracted representation than the commonly used metric maps, which show space to scale. For this reason, topological maps are more scalable, in that they do not require as much storage as equivalent metric maps. Further, topological maps are also more intuitive, and there are indications that topological representations are used by humans for navigation (Lynch 1971).

Owing to their abstractness, the construction of topological maps is arguably more complicated than the construction of metric maps. The problem involves the transformation of sensor measurements such as laser scans and odometry, most of which are metric in nature, into topological quantities. Further, the two component problems of topological detection, i.e. the first of determining which places are significant enough to be declared as nodes in the graph, and second of finding the correspondence between nodes of the topological graph and measurements, are hard problems in themselves. A systematic algorithm for the construction of topological maps has proved elusive.

In this paper, we address the correspondence problem, or the loop-closing problem, in the context of topological mapping. This involves matching sensor measurements to previously seen places, or being able to identify whether the measurement comes from a previously unseen place. These matches lead to identification of the global structure of the environment, which is nothing but the topological map. The problem is made difficult due to the fact that the same place may appear different at different times due to changes in lighting and viewpoint, among other things. At the same time, different places may also appear similar leading to *perceptual aliasing*. Current mapping algorithms mostly solve loop closing in an *ad-hoc* manner by matching low-level sensor-specific features. Even then, since exhaustive matching is intractable, heuristics such as nearest-neighbor matching are used, which make the solution even more brittle.

[1]Honda Research Institute USA, Mountain View, CA, USA
[2]College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

Corresponding author:
Ananth Ranganathan, Honda Research Institute USA, 425 National Avenue, Mountain View, CA 94043, USA.
Email: aranganathan@honda-ri.com

A systematic solution to the loop-closing problem in topological space needs to consider and evaluate all possible loop closures. This is because any solution that makes irreversible loop-closing decisions at any point in time can lead to a wrong solution, as a loop closure that appears correct now may be invalidated by future measurements. Hence, only by considering all possibilities can an algorithm gain the robustness that is not achievable through heuristics such as nearest-neighbor matching. This poses a computational problem since the space of all possible loop closings is combinatorial in nature and grows hyper-exponentially with the number of measurements.

We describe a topological mapping algorithm, called Online Probabilistic Topological Mapping (OPTM), first proposed by us in Ranganathan and Dellaert (2006). Our topological mapping algorithm is unique in that it is based on Bayesian inference in the space of all possible loop closings or, equivalently, all possible topologies. The algorithm is based on our previously presented Probabilistic Topological Maps (PTM) framework (Ranganathan et al. 2006), but improves it significantly with a new inference mechanism that enables online performance. This was impossible with the original PTM algorithm due to runtime constraints.

Inference in the space of all topologies is performed using an online Monte Carlo algorithm based on particle filtering to compute the Bayesian posterior. Such a sampling algorithm is necessary since the space of all topologies is combinatorial, as we will demonstrate, and exact inference is impossible. Particle filtering includes one measurement at a time into the inference procedure to update the posterior, and hence can be used in an online manner, which is of practical importance for use on a robot. Further, our Bayesian approach makes the inclusion of multiple types of sensors possible, as the posterior can be computed using all available measurements. A data-driven proposal algorithm is presented for making the algorithm efficient.

We extend our work in Ranganathan and Dellaert (2006) by introducing a novel appearance model for computing vision-based topological maps. The appearance model we use is the multivariate Polya distribution, a bag of words model. The model is based on scale-invariant feature transform (SIFT) features detected on images. We demonstrate the working of the OPTM algorithm using extensive experiments on a physical robot in diverse environments, including those not presented previously. The use of odometry, vision, and laser sensors attest to the sensor-independence of the framework. Since we are not concerned with landmark detection, the experiments involve manual landmark detection, as well as scenarios with landmarks placed at equidistant intervals and automatically detected landmarks, that validate the robustness of the algorithm. The experiments also include standard datasets that are popular for evaluation in the simultaneous localization and mapping (SLAM) community.

In the next section, we provide some background and related work followed by an exposition of the PTM framework. Subsequently, the OPTM algorithm is explained, including a description of the various sensor models used to integrate measurements. After this, some techniques for improving efficiency, such as the use of data-driven proposals, are presented followed by experiments and results.

2. Background and related work

PTMs were first introduced in Ranganathan and Dellaert (2004), where Markov chain Monte Carlo (MCMC) was used for inference in the space of topologies. The OPTM algorithm presented here was proposed in Ranganathan and Dellaert (2006), and this paper significantly improves upon that work by introducing a powerful appearance model and evaluating the system through new and extensive experiments.

In addition to topological mapping, our work relates to the area of loop closing in robotic mapping. In the following we review relevant prior research in these areas starting with topological mapping.

2.1. Topological mapping

A topological map consists of a graph containing nodes and edges. In this work, we use the standard definition of a topological map (Kuipers and Byun 1991; Choset and Nagatani 2001; Remolina and Kuipers 2004), wherein the nodes represent landmarks and the edges connecting them denote traversibility. The landmarks, in turn, are defined as distinct places in the environment, where the characteristics of the environment change significantly, such as at corridor junctions and doorways.

Topological mapping involves the important subproblem of deciding whether a landmark has been previously visited by the robot and, if so, when. This is also known as the *correspondence problem* in topological mapping. Solving the correspondence problem is made difficult due to perceptual aliasing in the environment, whereby many distinct landmarks appear to be similar to the robot's sensors. Thus, the robot has trouble labeling the landmark correctly and, consequently, in inferring the correct topology. Solving the correspondence problem has been a major area of research since solutions would also be applicable to the loop-closing problem in metric mapping (Thrun et al. 1998b).

Most existing techniques approach the mapping problem in a maximum-likelihood framework with the aim of finding the topology that minimizes some error function. The pioneering work in this regard is by Shatkay and Kaelbling (1997) that uses the Baum–Welch algorithm, a variant of the expectation–maximization (EM) algorithm used in the context of hidden Markov models (HMMs), to solve the aliasing problem for topological mapping. Other examples of HMM-based work include Kaelbling et al. (1996),

Gutierrez-Osuna and Luo (1996) and Aycard et al. (1997) where a second-order HMM is used to model the environment. A variation from the maximum-likelihood methods is the topological mapping system given by Goedemé et al. (2007) that uses image clustering to define regions of space as nodes in the topology. Loop closing and correspondence are performed using Dempster–Shafer decision theory, but again the decision is binding once taken. For all of the above methods that rely on picking the most likely topology, the most likely topology can frequently be wrong in the presence of aliasing. In addition, the error function to be optimized may have local minima which also results in an incorrect map.

Kuipers and Beeson (2002) apply a clustering algorithm to the measurements to identify distinctive places. This reduces perceptual variability but not perceptual aliasing, which is subsequently handled using logic-based hypothesis testing through further exploration. While this can work in many cases, no amount of exploration is guaranteed to eliminate ambiguity, nor does the method quantify ambiguity in usable terms.

Many existing algorithms use low-level characteristics specific to particular sensing modalities such as obstacle distances from laser scanners to characterize landmarks. These methods cannot be retargeted to other sensors. An example is Valgren et al. (2006) who perform topological mapping using an omnidirectional camera and model places using SIFT histograms. Ambiguity is solved using maximum-likelihood matching of SIFT features, by computing an affinity matrix of the images, and thus involves binding decisions at each step. Spectral clustering using the affinity matrix is also performed by Newman et al. (2006), albeit for the loop-closing problem in the context of metric maps. Cummins and Newman (2008) also provide an appearance-based technique for loop closing based on generative-modeling of bag-of-words models of location images. However, the decision of whether to close a loop is based on a maximum-likelihood computation. Dedeoglu et al. (1999) provide a mapping technique that uses specific features of the environment such as open doors and orthogonal walls, and identifies them using low-level characteristics of laser scans. Dudek and Jugessur (2000) use Fourier transforms of feature patches detected using attention operators for recognizing landmarks and overcoming ambiguity.

A common way of overcoming perceptual aliasing involves exploration by the robot until a distinct landmark is observed that localizes the robot. Examples of this approach include Choset's Generalized Voronoi Graphs (Choset and Nagatani 2001) and Kuipers' Spatial Semantic Hierarchy (Kuipers and Byun 1991). Other approaches that involve behavior-based control for exploration-based topological mapping are also fairly common. Matarić (1990) 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 sensorimotor control to topological and metric maps is described in Pierce and Kuipers (1997). Yamauchi and Beer (1996); Yamauchi and Langley (1997) 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. While the use of control is a valid approach, it can be wasteful in terms of time and energy. Our work, in contrast, attempts to extract the maximum information possible from available data, although it is also general enough to incorporate an active localization approach if needed.

Although more robust approaches that track multiple topological hypotheses when encountering ambiguity exist, these are limited in the sense that the whole space of hypotheses is not explored due to its combinatorial nature. For instance, Thrun et al. (1998b) 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. Another recent approach gives an algorithm to build a tree of all possible topological maps that conform to the measurements, but in a non-probabilistic manner (Remolina and Kuipers 2004; Savelli and Kuipers 2004). This has been integrated into the Hybrid Spatial Semantic Hierarchy (Hybrid SSH) model that integrates metric and topological mapping at different scales to provide a comprehensive mapping framework (Kuipers et al. 2004). Dudek et al. (1993) have also given a technique that maintains multiple hypotheses regarding the topological structure of the environment in the form of an exploration tree.

An approach that is closer to our ideal in the sense of maintaining a multi-hypothesis space over correspondences, is given by Tomatis et al. (2002) and also uses partially observable Markov decision processes (POMDPs) to solve the correspondence problem. However, in their case the multi-hypothesis space is used only to detect the points where the probability mass splits into two. Also, like a lot of others, this work uses specific qualities of the indoor environment such as doors and corridor junctions, and hence is not generally applicable to any environment. Similarly, Tapus (2005) proposes the use of POMDPs for disambiguation. The distinguishing features of this work is however, the use of 'fingerprints of places' that incorporate various different features such as edges, lines, and color histograms, and help in resolving ambiguity to a significant extent. Work by Modayil et al. (2004) generates an ensemble of topological maps and uses them to construct a global metric map. However, they do not provide a probabilistic ordering to their ensemble of maps as the posterior on topologies constructed by our algorithm does.

Recent work by Blanco et al. (2008) describes a comprehensive method called HMT-SLAM for constructing hybrid metric–topological maps wherein the global map structure is topological, but is represented locally using metric submaps. The formalism for computing the loop-closing

hypothesis in the topology is similar to ours and is borrowed from our previous work in Ranganathan and Dellaert (2006). Normalized cuts, based on the partition of measurements by the topology, are used to compute the likelihood of each topology (Blanco et al. 2006).

2.2. Loop closing

Loop closing in metric maps implies identifying correspondences between measurements and physical locations. This problem in its general form is combinatorial and, hence, most solutions use heuristics to reduce the search space. However, reverting to a specific correspondence assignment using any heuristic involves the danger of choosing a wrong correspondence, often leading to catastrophic failure of the mapping system. Common mapping approaches assume known correspondences, or use a maximum-likelihood assignment, for instance using the EM algorithm (Thrun et al. 1998a; Burgard et al. 1999). Maximum-likelihood data association using the uncertainty covariance on the robot poses to eliminate unlikely matches is also common (Gutmann and Konolige 2000). Pradeep et al. (2009) describe a method similar to Blanco et al. (2008) in that a hybrid metric–topological map is built. Bundle adjustment is used to obtain the local metric submaps while a loop closing test based on feature matching is used to get the global topology.

A robust way of dealing with unknown correspondences is to generate a set of candidate assignments, and then prune inconsistent ones. This is typically done using random sample consensus (RANSAC) by Bolles and Fischler (1981), a probabilistic algorithm that repeatedly selects a minimum number of candidates needed to constrain the given problem, and determines the support from the remaining candidates. It is frequently used in visual SLAM (Se et al. 2005; Williams et al. 2007).

A more direct way of dealing with unknown correspondences is to maintain multiple likely hypotheses simultaneously. This provides a computational compromise between picking the most likely solution and dealing with the complete space of loop closings. Particle filters, where each particle maintains a map based on its own set of correspondences, have been popular (Eliazar and Parr 2003; Hähnel et al. 2003a; Montemerlo and Thrun 2003), in spite of the use of importance sampling in such high-dimensional continuous spaces being theoretically unjustified, and practically problematic. The joint compatibility branch and bound algorithm by Neira and Tardos (2001) has also proven successful (Thrun et al. 2005; Bailey and Durrant-Whyte 2006).

A systematic approach, closely related to our own, that avoids the combinatorial nature of the problem samples from probability distributions over correspondences (Dellaert 2001; Dellaert et al. 2003) in a batch structure from motion context, and has also been adapted to SLAM (Kaess and Dellaert 2005). However, these methods are computationally quite expensive. A recent method that uses a simple but fast bag-of-words model for vision-based loop closing is that of Angeli et al. (2008). The probability of loop closing is maintained and a high probability location is checked through epipolar geometry, which can be computationally expensive. Moreover, loop-closing decision once taken cannot be reversed and, hence, the management of loop closing itself is not probabilistic.

3. Probabilistic topological maps

We begin by giving an overview of the PTM framework, which solves the aliasing problem in a systematic manner by computing the posterior over the space of topologies. 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, PTMs provides a quantification of uncertainty by pegging a probability of correctness to each topology. On the other hand, current techniques simply return a wrong topological map without any indication of error or ambiguity.

However, before we can perform inference in the space of topologies, we need to understand the nature of this space. Consider a scenario where a robot moves around an environment and visits six locations that are deemed to be landmarks. It is required that the robot identify the topology of the environment from these six observations. Consider two specific cases (note that these are not the only two possible): one in which each of the landmarks is unique and the second in which the second and the last measurements come from the same landmark. We can illustrate these two scenarios as shown in Figure 1. It can be seen that the measurements corresponding to the same landmark can be grouped into a set, and this grouping then defines a set partition on the set of measurements. Each set partition, in turn, is equivalent to a topology of the environment. As an additional example, the space of topologies containing 15 possible topologies for the case of four measurements is shown in Figure 2.

The key idea behind inference in the space of topologies is this equivalence between topologies of an environment and set partitions of landmark measurements. The set partition corresponding to the topology can also be viewed as a label sequence as shown in Figure 3. Each set in the partition corresponds to a distinct label, which in turn, corresponds to a unique landmark. More formally, associating a label with each landmark, we can represent the topology by a label sequence $L^n = L_{1:n}$, where L_i is the label of the ith landmark. Further the number of unique labels in this sequence is equal to the number of observed unique landmarks in the environments. The posterior on the space of topologies that we seek can then be written as $p(L^n|Z)$.

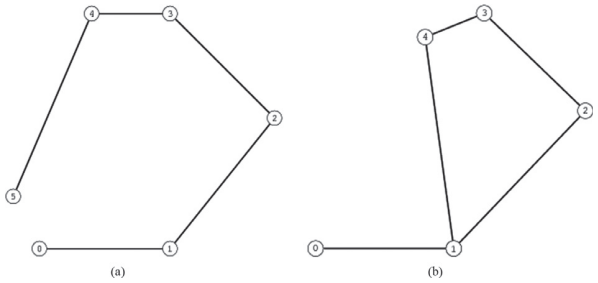


Fig. 1. Two topologies with six observations, each corresponding to set partitions (a) with six landmarks ($\{0\}, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}$) and (b) with five landmarks ($\{0\}, \{1, 5\}, \{2\}, \{3\}, \{4\}$), illustrate the equivalence between topologies and set partitions.

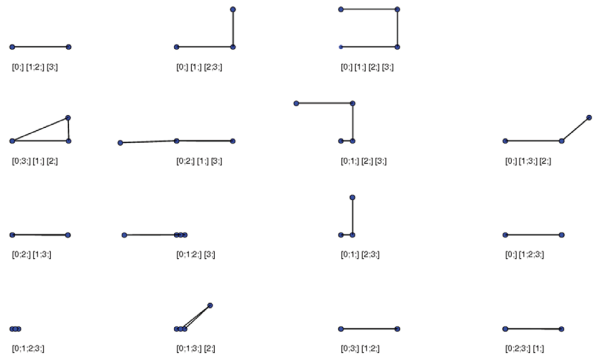


Fig. 2. There are 15 possible topologies for the case of four measurements. The set partitions corresponding to the topologies are given below each topology.

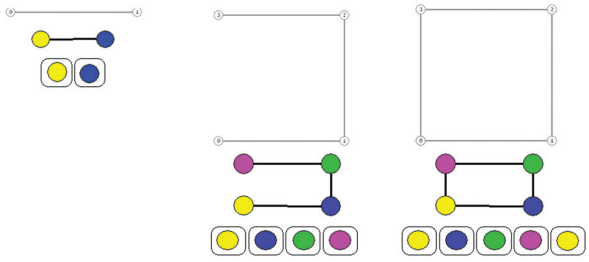


Fig. 3. An example of topologies as label sequences (bottom), with each label shaded (colored) differently. Each label sequence corresponds to a set partition.

It can be seen that the labels induce a set partition of the measurements, which results in the above-mentioned isomorphism between topologies and set partitions. The number of possible topologies for a given environment is thus equal to the number of set partitions of the set of measurements. This number is called the *Bell number* (Nijenhuis and Wilf 1978), and grows hyper-exponentially with the number of measurements.

A PTM is simply the posterior on the space of topologies. It is thus a set of topologies with their associated posterior probabilities. However, naive computation of the PTM by simple enumeration of the topologies is not possible due to the combinatorial nature of the space. Instead, we compute a sample-based approximation of the posterior. This is possible since, in almost all practical scenarios, most of the topologies have a posterior probability close to zero and need not be evaluated. Only a small set of similar topologies that agree with the measurements have non-negligible probabilities, as shown in Figure 4, and can be represented efficiently using sample-based inference.

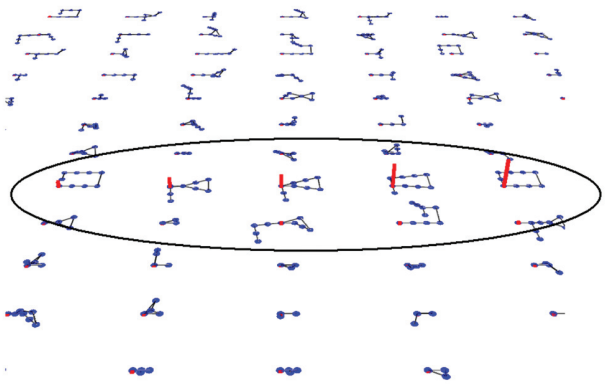


Fig. 4. Even though the space of topologies is combinatorial, topologies with non-negligible probabilities are relatively few and localized. A portion of the space of topologies obtained with ten observed landmarks is shown. The groundtruth is the topology with the highest probability

The above assumption is violated only when the environment is highly perceptually aliased. In these cases, the algorithm is no longer efficient but is still capable of quantifying the ambiguity since many topologies will have almost similar probabilities in the posterior. Existing algorithms are incapable of this and fail silently in such cases.

Sample-based inference using particle filtering is explained in the next section.

4. Online topological mapping using Rao–Blackwellized particle filters

We now describe the OPTM algorithm. For ease of exposition of the algorithm, a summary of all of the notation used is given in Table 1. To reiterate the conclusion of the discussion above, the posterior on topologies that we seek is the distribution of label sequences $P(L^n|z^n, a^n)$.

In our case, the measurements consist of appearance measurements derived from camera images $a^n = \{a_1, a_2, \ldots, a_n\}$, laser range scans $s^n = \{s_1, s_2, \ldots, s_n\}$, and odometry measurements between landmarks $o^n = \{o_1, o_2, \ldots, o_{n-1}\}$. The posterior on topologies that we seek is represented as $p(L^n|a^n, s^n, o^n)$. Applying Bayes law on the required posterior to obtain the measurement likelihood and prior, we obtain

$$p(L^n|z^n, a^n) \propto p(L^n|z^{n-1})\, p(a^n, s_n, o_{n-1}|L^n, z^{n-1}), \quad (1)$$

Table 1. Notation used in the explanation of the algorithm.

Symbol	*Meaning*
n	Total number of landmarks observed
m	Number of distinct landmarks observed
o^n	The $n-1$ odometry measurements
s^n	Range scan measurements around the n landmarks
a^n	Appearance measurements from the n landmarks
z^n	Combined set of metric measurements $z^n = \{s^n, o^n\}$
L^n	Topology represented as a label sequence
X^n	Landmark locations for the topology L^n
$\alpha_n(X^n)$	Analytic distribution on the landmark locations

where the metric measurements up to the nth landmark have been represented as $z^n = \{s^n, o^n\}$, and the likelihood of the measurements from the nth observed landmark is $p(a^n, s_n, o_n|L^n, z^{n-1})$. The appearance measurement likelihood is evaluated in batch due to the nature of the appearance model as will be shown in Section 4.3. The prior $p(L^n|z^{n-1})$ can be further factorized to give an incremental prior on the label in the current (nth) time step

$$p(L^n|z^{n-1}) = p(L_n|L^{n-1}, z^{n-1})\, p(L^{n-1}|z^{n-1}), \qquad (2)$$

where $p(L_n|L^{n-1}, z^{n-1})$ is the prior (proposal) distribution for the label on the nth observed landmark and $p(L^{n-1}|z^{n-1})$ is the posterior from the previous step containing $n-1$ measurements. The prior gives a distribution on which of the distinct landmarks we are likely to see next, including the possibility of the next landmark being a previously unvisited one. It can be seen that Equations (1) and (2) together give a recursive formulation for the posterior on topologies that is amenable for performing particle filtering.

Evaluation of the odometry and laser likelihoods can be done independently of the appearance likelihood, and requires metric information since these measurements relate to metric quantities in the environment. In this case, the sufficient statistic in metric space for evaluating these likelihoods is the location of the landmarks. Hence, we marginalize over the landmark locations to compute the likelihood

$$p(s_n, o_n|L^n, z^{n-1}) = \int_{X^n} p(s_n, o_{n-1}|L^n, X^n, z^{n-1}) \times p(X^n|L^n, z^{n-1}), \qquad (3)$$

where X^n is the vector of landmark locations of length n and we have used the chain rule in the integrand. Note that the prior on landmark locations $p(X^n|L^n, z^{n-1})$ can be further factorized into a predictive prior on the location of the current (nth) landmark and the posterior on locations from the previous step

$$p(X^n|L^n, z^{n-1}) = p(X_n|L^n, X^{n-1}, z^{n-1}) \times p(X^{n-1}|L^{n-1}, z^{n-1}), \qquad (4)$$

where X_n is the location of the nth landmark and $p(X^{n-1}|L^{n-1}, z^{n-1})$ is the posterior on landmark locations from the previous step. Furthermore, the integrand of (3) is equal to the posterior on X^n up to a normalization constant. Hence, we also have a recursive formulation for the posterior on landmark locations $p(X^n|L^n, s^n, o^{n-1})$.

Since storing the posterior on landmark locations at each step aides in the computation of the measurement likelihood, we add this information to each of the particles. However, this is a large continuous space and joint sampling of this space with the space of topologies is not possible. Instead, the posterior is stored in an analytical form, a Gaussian distribution in our case, and updated at each step.

A Rao–Blackwellized particle filter (RBPF) maintains the joint posterior over two disparate spaces by representing part of it analytically and sampling over the remaining part of the space. Hence, our motivation for using a RBPF in our algorithm is clear. The RBPF maintains the posterior $p(L^n, X^n|s^n, o^n)$ on the joint space of landmark locations and topologies but in a hybrid discrete-continuous form. This posterior can be denoted by a set of hybrid weighted samples containing a topology and an analytical marginal posterior on the landmark locations conditioned on the sample value

$$S_n = \left\{L^{n,(i)}, w_n^{(i)}, \alpha_n(X^n)\right\}_{i=1}^{N}, \qquad (5)$$

where $w_n^{(i)}$ is the weight on the ith particle and $\alpha_n^{(i)}(X^n) \triangleq p(X^n|L^{n,(i)}, s^n, o^n)$ is the analytic form of the landmark location posterior. An example of a joint sample from the RBPF is shown in Figure 5.

The two components required to perform filtering are the proposal distribution and a method for computing the importance weights. These are explained in the following sections.

4.1. The proposal distribution

We use the predictive prior distribution on the current landmark label $p(L_n|L^{n-1}, z^{n-1})$, given in Equation (2), as our proposal distribution. Using the sample notation of Equation (5), the proposal distribution can be written as

$$L_n^{(i)} \sim p\left(L_n|L^{n-1,(i)}, z^{n-1}\right). \qquad (6)$$

This is a discrete probability distribution on a vector of size $p+1$, where p is the number of distinct landmarks observed up to the $(n-1)$th step. The distribution (6) encodes our expectation of the robot revisiting one of the previously observed landmarks or visiting a completely new one. While a uniform distribution can be used for this purpose, it does not capture all of the characteristics of the problem. For example, with a uniform distribution, the probability of a robot visiting a new landmark remains constant with time. However, we can reasonably expect the robot to visit fewer new landmarks as its run progresses. Similarly, landmarks that have been visited frequently in the past should be better candidates for revisitations. This is especially true for indoor environments where lobbies and corridor junctions are visited more frequently than other locations.

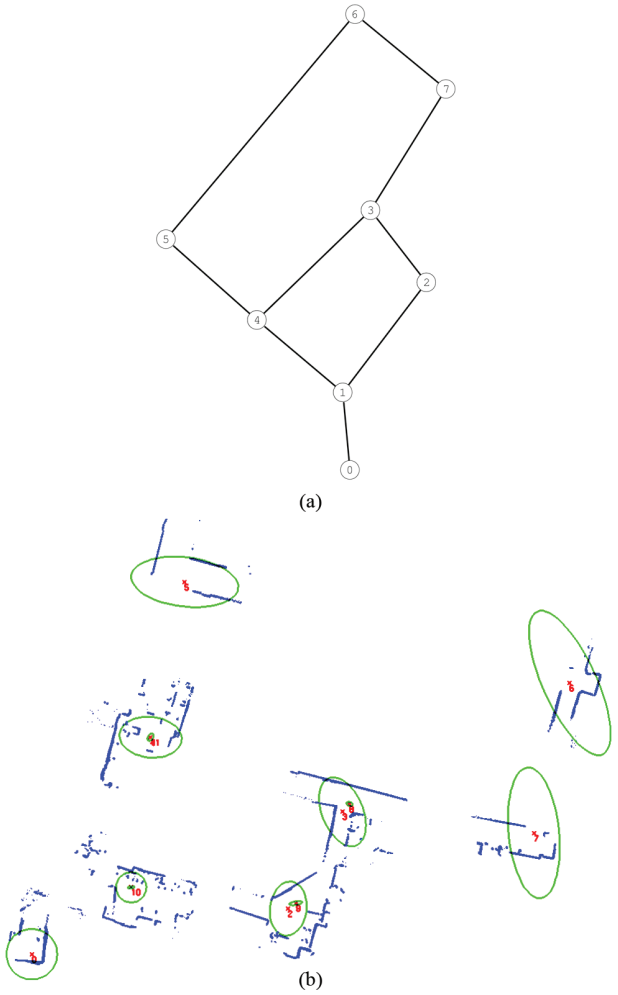


Fig. 5. A sample from the RBPF that contains (a) a topology and (b) an analytical distribution on the landmark locations in the form of a Gaussian. The points (denoted by crosses) in (b) are the mean landmark locations while the ellipses denote marginal covariances.

A distribution that models these problem characteristics well is the Dirichlet process prior (Blackwell and MacQueen 1973; Ferguson 1973). The Dirichlet process is an extension of the standard Dirichlet distribution to infinite mixture models, i.e. it also includes a probability of observing previously unobserved measurement classes, in our case landmarks. The prior on the nth landmark label using the Dirichlet process is given as

$$p\left(L_n^{(i)}|L^{n-1,(i)}\right) = \begin{cases} \frac{n(L_j^{(i)})}{n+c} & 1 \leq j \leq p, \\ \frac{c}{n+c} & j = p+1, \end{cases} \tag{7}$$

where p is the number of distinct landmarks observed up to the $(n-1)$th step as before, and $n(L_j^{(i)})$ is the number of occurrences of the label j in the label sequence corresponding to the topology.

The parameter c encodes our belief in the number of distinct landmarks in the environment. A large value of c increases the probability of observing a new landmark at every step and, consequently, the number of distinct landmarks in the topology. Note that the probability of observing a new landmark decreases as n increases, although it never goes down to zero. Also, the probability of revisiting a landmark is proportional to the number of times it has been visited before, given by $n(L)$.

4.2. Importance weight computation

The importance sampling weights for the particle filter are defined as ratio of the target distribution to the proposal distribution, so that

$$w_n^{(i)} = \frac{\text{Target distribution}}{\text{Proposal distribution}} w_{n-1}^{(i)} \tag{8}$$

$$\propto p(a^n, s_n, o_n|L^{n,(i)}, z^{n-1})\, w_{n-1}^{(i)} \tag{9}$$

where we have used the target distribution from (1) and proposal from (2).

The appearance measurement is assumed conditionally independent of the scan and odometry measurement given the topology so that measurement likelihood can be written as

$$p(a^n, s_n, o_n|L^{n,(i)}, z^{n-1}) = p(a^n|L^{n,(i)}, z^{n-1}) \times p(s_n, o_n|L^{n,(i)}, z^{n-1}). \tag{10}$$

A bag-of-words model based on SIFT features is used for the appearance measurements. For the laser measurements, range scans are used to construct local map patches around landmark locations that the robot visits. These map patches are subsequently matched using scan matching techniques to provide a likelihood of their being from the same physical location. This gives us a sensor model for laser scans obtained at the landmark locations. We describe the evaluation of these measurement likelihoods in the following sections

4.3. Appearance modeling using 'bag-of-words' models

Recently, SIFT (Lowe 2004) descriptors of features have gained greatly in popularity due to their ability to convert many types of features into a standard, reproducible vector space. Furthermore, SIFT descriptors of features detected on an image can be quantized and the quantized features can be considered to be the analogue of words in a document. Thus, the image becomes a sequence of 'appearance words', making document analysis methods applicable to images. In keeping with the text processing community, methods that model images by converting SIFT features into appearance words are called 'bag-of-words' models (Sivic et al. 2005) since they do not consider the sequence of the words themselves, but only their occurrence frequency.

Appearance words are learned in an offline phase wherein the 128-dimensional SIFT space is vector quantized. SIFT features extracted from training images

obtained from the environment are used for this purpose. The number of bins in the vector quantization, which corresponds to the number of words in a text document, is a parameter. Vector quantization is performed using the K-means algorithm, which clusters the features detected on the training image set. During runtime, each image is transformed into a histogram of word counts in each of the bins. Thus, the representation of an image in a bag-of-words model is a vector of word counts, which comprise a histogram. In practice, we use a set of images obtained from the test environment for learning the visual words, though any other images not too dissimilar from the test environment could be used.

4.4. The multivariate Polya model

The aim of using the image histograms is to identify landmarks that are physically the same. However, since we have an appearance model that is conditioned on the topology, all that is needed is to evaluate the correspondence that is implied by the topology.

We model all of the images arising from a landmark as having the same underlying global characteristics. Since the measurements are histograms of word counts, they are modeled using a multinomial distribution having its dimensions equal to the number of appearance words. Further the prior over the multinomial parameter is the conjugate Dirichlet distribution (not to be confused with the Dirichlet process introduced above) to aid in ease of computation

$$P(a^n|L^n) = \prod_{\xi \in L^n} \int_{\theta_\xi} P(\theta_\xi|\alpha_\xi) \prod_{a \in \xi} P(a|\theta_\xi), \qquad (11)$$

where $\{a\}$ is the set of measurements indexed by set ξ, and $\theta_\xi = [\theta_{\xi 1}, \theta_{\xi 2}, \ldots, \theta_{\xi W}]$ and $\alpha_\xi = [\alpha_{\xi 1}, \alpha_{\xi 2}, \ldots, \alpha_{\xi W}]$ are the multinomial parameter and Dirichlet prior, respectively. Note that ξ indexes the sets in the set partition corresponding to the topology L^n as was explained in Section 3. The number of distinct appearance words is denoted as W. Let the total number of features in an image be denoted N and the number of features in each visual word be denoted as $N_1, N_2, \ldots, N_W$. Then, the distributions in the integrand above are

$$p(a|\theta_\xi) = \frac{N!}{N_1! N_2! \ldots N_W!} \theta_{\xi 1}^{n_1} \theta_{\xi 2}^{n_2} \ldots \theta_{\xi W}^{n_W}, \qquad (12)$$

$$p(\theta_\xi|\alpha_\xi) = \frac{\Gamma(\sum_{w=1}^W \alpha_{\xi w})}{\sum_{w=1}^W \Gamma(\alpha_{\xi w})} \theta_{\xi 1}^{\alpha_{\xi 1}-1} \theta_{\xi 2}^{\alpha_{\xi 2}-1} \ldots \theta_{\xi W}^{\alpha_{\xi W}-1}. \qquad (13)$$

The likelihood model in (11) with the multinomial and Dirichlet distributions defined in (12) and (13) is called the multivariate Polya model, or equivalently in document modeling, the Dirichlet compound multinomial model (Madsen et al. 2005). The multivariate Polya model models burstiness in the data, i.e. the empirical observation that if a word occurs once in a document, it is likely to occur many more times. However, as with any bag-of-words model, it assumes that the occurrence of different words is independent of each other.

The integration in (11) can be performed in closed form since the Dirichlet distribution is the conjugate prior of the multinomial distribution. This yields the final form of the appearance likelihood as

$$P(a^n|L^n) = \prod_{\xi \in L^n} \frac{N^{(\xi)}!}{\prod_{w=1}^W N_w^{(\xi)}} \frac{\Gamma(\alpha_\xi)}{\Gamma(N^{(\xi)} + \alpha_\xi)} \prod_{w=1}^W \frac{\Gamma(N_w^{(\xi)} + \alpha_{\xi w})}{\Gamma(\alpha_{\xi w})}, \qquad (14)$$

where $N_w^{(\xi)}$ is the total count of the wth appearance word across all of the image histograms in set ξ and $N^{(\xi)} = \sum_w N_w^{(\xi)}$, $\alpha_\xi = \sum_w \alpha_{\xi w}$.

Given a collection of D images with features detected on them, the maximum-likelihood value for α can be learned by using iterative gradient descent optimization. It can be shown that this leads to the following fixed point update (Minka 2003)

$$\alpha_w^{\text{new}} = \alpha_w \frac{\sum_{d=1}^D \psi(n_{dw} + \alpha_w) - \psi(\alpha_w)}{\sum_{d=1}^D \psi(n_{dw} + \alpha) - \psi(\alpha)}, \qquad (15)$$

where $\alpha = \sum_w \alpha_w$ as before, and $\psi(\cdot)$ is the digamma function (derivative of the Gamma function).

For a new landmark, the number of images in the set ξ is one, and it is not possible to learn a coherent maximum likelihood using (15) above. We overcome this by learning a more 'general' prior value for α from a training set of images. We use the same set of images for this purpose as was used for learning the visual words. Owing to the broad prior, new landmarks always receive a low appearance likelihood.

The appearance likelihood, evaluated by learning the α parameter for each set in the topology and using these values appropriately in (14), can be used in the particle filtering algorithm for computing the importance weight in (10).

4.5. Odometry and laser scan likelihood evaluation

Odometry and laser scan likelihoods are evaluated by introducing the landmark locations and marginalizing over them as given in (3). Evaluating (3), in turn, requires the computation of the conditional likelihoods on the scan and odometry measurements. Assuming the independence of the scan and odometry measurements given the landmark measurements, we obtain

$$p(s_n, o_n|L^{n,(i)}, X^n, z^{n-1}) = p(s_n|L^{n,(i)}, X^n, z^{n-1}) \times p(o_n|L^{n,(i)}, X^n, z^{n-1}). \qquad (16)$$

The scan likelihood $p(s_n|L^{n,(i)}, X^n, z^{n-1})$ is obtained by performing scan matching between the map patches from

Fig. 6. Scan measurements, obtained by concatenating scans from around landmark locations, used by the RBPF algorithm.

the landmark locations. The map patches are obtained, in turn, by simply concatenating laser scans from a local area around the landmark as the robot moves through it (Figure 6). We use the scheme of Chen and Medioni (1991), which involves point-to-plane matching, to perform scan matching. Since odometry drift can affect this simple scheme for creating the map patches, we limit the size of the patch to a 5 meter robot trajectory length. The odometry likelihood $p(o_n|L^{n,(i)}, X^n, z^{n-1})$ is evaluated simply through the use of an odometry model.

The prior on the landmark location $p(X_n|L^{n,(i)}, X^{n-1}, z^{n-1})$ from (4) encodes the notion that distinct landmarks do not usually occur close together in the environment. We use the same prior on landmark locations as given in Ranganathan et al. (2006). Topologies which place distinct landmarks close together in location are penalized by this prior.

The scan likelihood for a new landmark can be computed in exactly the same way as described above since the map patch for a landmark location is generated immediately. However, in contrast to the appearance case, scan matching the measurement scan against this map patch results in a very high likelihood since the scan will match perfectly. However, this high laser likelihood for a new landmark is counteracted by the landmark prior, odometry measurements and also by the appearance likelihood. This absolves us of the need to specify an elaborate likelihood evaluation scheme for the case of a new landmark as is done in Blanco et al. (2008) and Angeli et al. (2008).

Note that the scan likelihood evaluation described above only works since in our case a measurement is only obtained at a landmark location, i.e. when the landmark detector module fires. As this does not occur frequently, the above approach works when odometry is available, even if appearance information is ambiguous. However, when odometry is also absent or ambiguous, our approach may break down as the algorithm is forced to rely on the prior alone to correct for the high likelihood assigned by the laser measurement to new landmarks.

4.6. Summary of the OPTM Algorithm

As the odometry model is assumed to be Gaussian and the result of the scan matching operation is also a Gaussian distribution, all of the distributions involved in the computation of the landmark location posterior (4) are Gaussian except for the landmark prior. This makes the posterior non-Gaussian.

The computation is kept recursive by projecting the non-Gaussian posterior onto a Gaussian posterior using the Laplace approximation. This involves replacing the true posterior by a Gaussian centered around the *maximum a posteriori* (MAP) value of landmark locations. In practice, this step is performed by linearizing around the most likely landmark location, which is found through an optimization. Details can be found in Ranganathan et al. (2006). The covariance at the MAP location is estimated through the Hessian matrix obtained from the optimization algorithm. The OPTM algorithm thus trades off maximum-likelihood estimation, or in this case really MAP estimation, in the continuous landmark space with a much richer approximation of the probability distribution in the discrete space of topologies, which is really the object of our interest.

The weight computation of (9) can now be performed in closed form as it involves integrating a Gaussian distribution, albeit unnormalized.

Algorithm 1 The OPTM algorithm.

1. Randomly select a particle $L^{n-1,(i)}$ from the previous time step according to the weights $w_{n-1}^{(i)}$.
2. Propose a new topology sample using the proposal distribution $p\left(L_n^{(j)}|L^{n-1,(i)}\right)$ in (6)
3. Calculate the Gaussian posterior density on landmark locations $\alpha_n^{(j)}(X^n)$ using Bayes law as in (4) and the Laplace approximation.
4. Calculate the importance weights $w_n^{(j)}$ from (10). The appearance likelihood is calculated using (14), and the odometry and scan likelihoods as the integral over the unnormalized $\alpha_n^{(j)}(X^n)$ in (3) and (16).

We now have all of the components to perform the inference using the RBPF. A summary of the OPTM algorithm is provided in Algorithm 1.

5. Data-driven proposals for particle filters

Particle filters are susceptible, over time, to the 'particle degeneracy' or 'lack of diversity' problem (Doucet et al. 2000), wherein only one sample has a significantly non-zero weight. The reason for this is that many samples fall into

regions of low probability and die out during the filtering process. This results not only in the failure to converge to the correct posterior but also in wasted computation, since the algorithm is evaluating the weights of samples that will be lost in any case.

A data-driven proposal overcomes this problem by proposing more samples from regions of high probability so that samples and computation are not wasted. Note that the proposal distribution in (6) does not make use of the current measurement. We rectify the situation in this section by presenting a proposal distribution that uses the odometry to provide more likely samples.

The key idea behind the data-driven proposal is that the odometry likelihood can be incorporated into the proposal distribution while only the appearance and scan likelihoods are used to compute the importance weights. The measurement likelihood in (1) is thus split into two parts. This split also entails a two-step process for updating the analytic posterior on landmark locations since this posterior needs to be updated using both the odometry and scan measurements.

In the following exposition, we do not consider appearance measurements since the appearance likelihood is evaluated in exactly the same manner and is unaffected by the data-driven proposal. Starting with the posterior on topologies, we factor it using Bayes law in a slightly different manner to obtain

$$p(L^n|s^n, o^n) \propto p(L^n|z^{n-1}, o_{n-1})\, p(s_n|L^n, z^{n-1}, o_{n-1}), \quad (17)$$

where the likelihood is factored into two terms using the chain rule. The prior term can in turn be written using Bayes law as the product of the odometry likelihood and a prior on the current label

$$\begin{aligned} p(L^n|z^{n-1}, o_{n-1}) \propto & p(L_n|L^{n-1}, z^{n-1})\, p(L^{n-1}|z^{n-1}) \\ & \times p(o_{n-1}|L^n, z^{n-1}). \end{aligned} \quad (18)$$

The proposal distribution is taken to be the right-hand side of (18), which can be written using the sample representation of (5) as

$$L_n^{(i)} \sim p\left(L_n|L^{n-1,(i)}, z^{n-1}\right) p(o_n|L^{n,(i)}, z^{n-1}). \quad (19)$$

The form of the predictive label distribution $p(L^n|z^{n-1})$ is the Dirichlet process prior as before. However, the odometry likelihood in (18) is evaluated by marginalization over the landmark locations

$$\begin{aligned} p(o_n|L^n, z^{n-1}) = & \int_{X^n} p(o_n|L^{n,(i)}, X^n, z^{n-1}) \\ & \times p(X^n|L^{n,(i)}, z^{n-1}), \end{aligned} \quad (20)$$

where the same landmark prior and odometry model are used as in Section 4.5. Note that the prior in (20) can be evaluated using the posterior on the landmark locations from the previous by use of the chain rule as in (4).

One drawback of this proposal distribution is the need to perform m optimizations to compute it. These optimizations are required since the integral in (20), evaluated by linearizing around the optimum, needs to be computed for all of the possible label values for L_n (except for the case when L_n is a new landmark), which are m in total. However, performing these extra optimizations once per filtering step is still preferable to evaluating the importance weight for all of the particles that do not survive when a vanilla proposal is used.

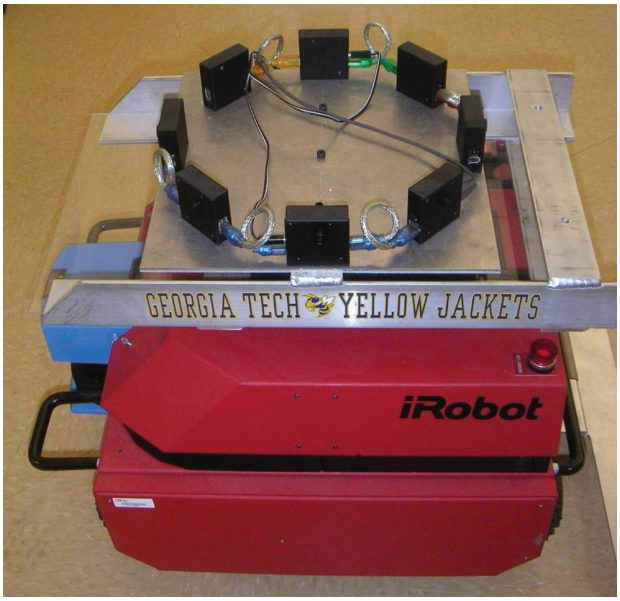


Fig. 7. ATRV-mini robot with eight camera rig and forward-facing laser scanner used in the experiments.

From the target (17) and proposal (18) distributions and the definition of the importance weights (8), we get the expression for the importance weights in this case as

$$w_n^{(i)} \propto p(s_n|L^{n,(i)}, z^{n-1}, o_n)\, w_{n-1}^{(i-1)}.$$

This is evaluated by marginalization over landmark locations

$$\begin{aligned} p(s_n|L^{n,(i)}, z^{n-1}, o_n) \propto & \int_{X^n} p(s_n|L^{n,(i)}, X^n, z^{n-1}, o_n) \\ & \times p(X^n|L^{n,(i)}, z^{n-1}, o_n), \end{aligned}$$

where the scan likelihood is evaluated using scan matching exactly as in Section 4.5, since it is independent of the odometry given the landmark locations. The location prior $p(X^n|L^{n,(i)}, z^{n-1}, o_{n-1})$ is the same as the integrand of (20) up to a normalizing constant and the linearized Gaussian approximation found therein is used again here.

6. Results

We now present the results of testing the OPTM algorithm in various environments. An ATRV-mini robot with an eight camera rig mounted on top and a laser scanner facing forwards, as shown in Figure 7, was used for the experiments. Particle filtering was performed using the optimal stratified resampling method of Fearnhead and Clifford (2003)

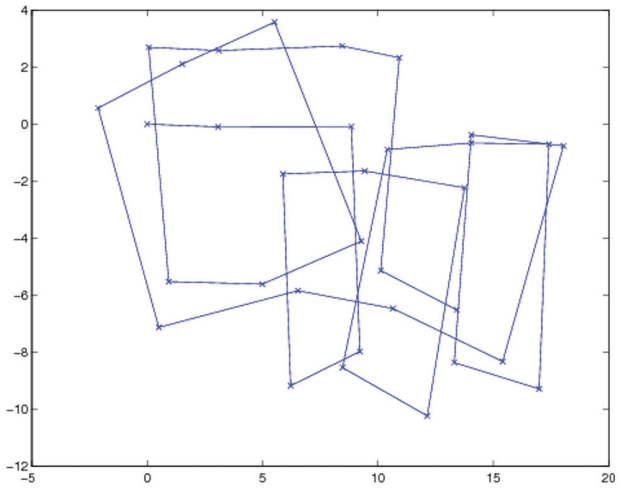


Fig. 8. Landmark locations obtained from simulated odometry for the first experiment.

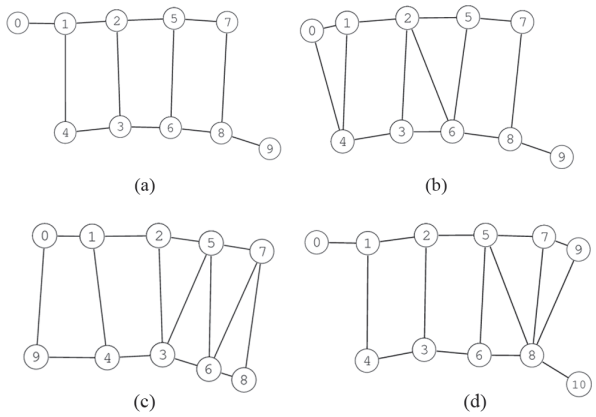


Fig. 9. Topologies with highest posterior probability mass for the second experiment. (a) The groundtruth topology receives 59% of the probability mass while (b), (c), and (d) receive 9.1%, 8.2%, and 6% of the probability mass, respectively.

that varies the number of particles automatically but ensures that memory usage is $O(1)$ with respect of number of measurements seen so far. The data-driven proposal was used in all of the experiments. A value of 3.0 was empirically determined for the Dirichlet prior parameter c. This value gives good results on all of our experiments, which have a widely varying number of landmarks. Changing the value of the prior parameter by up to a factor of 10 did not affect the results significantly. The SIFT-based appearance model described in Section 4.3 above was used to model images obtained from the camera rig. SIFT features detected from the landmark images were quantized into 1,024 appearance words. The landmark location prior was used with a value of 10 meters for the penalty radius and 15 for the maximum penalty value. For a description of these parameters and their effect on the inferred posterior, see Ranganathan et al. (2006). In the absence of an automatic landmark detector, landmarks were identified manually and corresponded to gateways and junctions in the environment (Schroter et al. 2003).

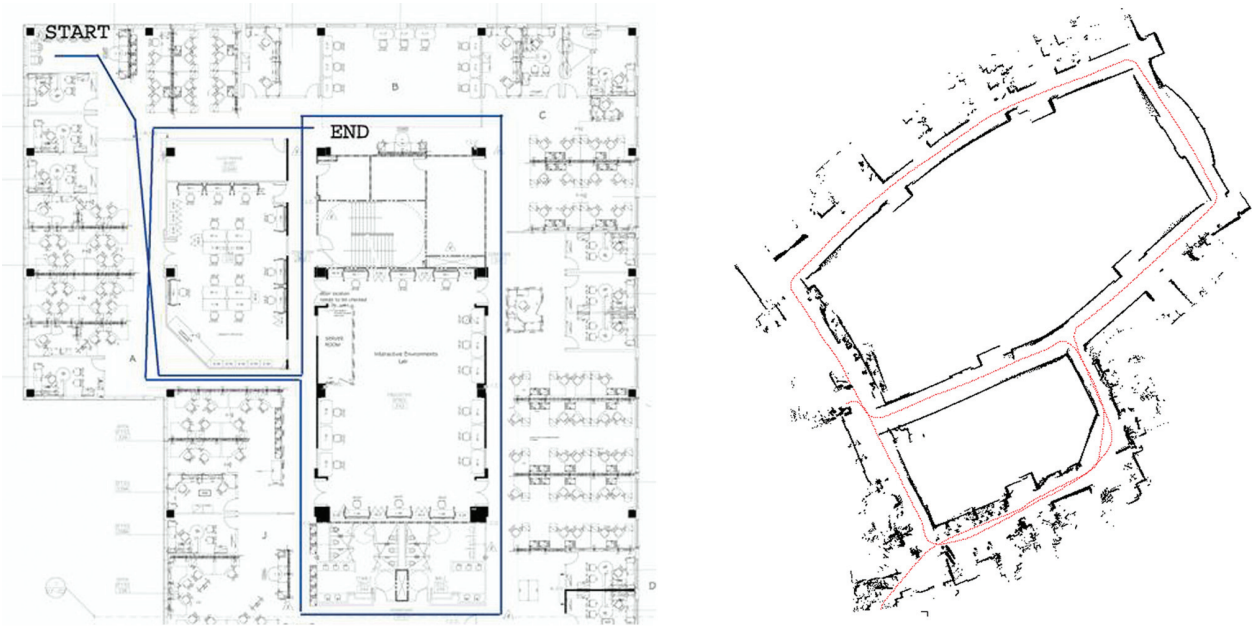
The first experiment was conducted using simulated data to demonstrate the capability of the OPTM to work with just odometry measurements similar to the PTM (Ranganathan et al. 2006). A very loopy sequence of length approximately 80 meters with 813 odometry measurements was generated and 33 landmarks were manually labeled in the trajectory. The simulated odometry with noise is shown in Figure 8 and the resulting posterior from the OPTM algorithm is shown in Figure 9. Owing to the absence of any laser and appearance measurements, the probability mass on the groundtruth topology is relatively low.

The second experiment was conducted in the TSRB building at Georgia Tech using the ATRV-mini robot in an indoor setting. A map of the experiment area along with the robot path, which is approximately 100 meters long and passes through 12 landmark locations, is shown in Figure 10(a). The map patches obtained by concatenating scans around the landmark locations are shown in Figure 6. The result of the OPTM algorithm is a joint distribution on topologies and landmark locations. The maximum-likelihood sample is shown in Figure 5. The distribution on the landmark locations is displayed in the figure through the marginal covariance ellipses along with the local map patches aligned using scan matching. The corresponding topology, shown in Figure 5(a), is also the groundtruth topology and obtains 98.5% of the probability mass in the posterior. The topology constraints and the inferred landmark locations can be used to produce a global metric map using the global optimization technique of Lu and Milios (1997). The resultant metric map is given in Figure 10(b). It can be seen that this simple post-processing step produces a globally consistent metric map.

A third experiment was performed in a larger environment (about 60 meters across) of the CRB building to confirm our findings. The RBPF computes the PTM that gives the groundtruth topology in Figure 11(a), 94% of the probability mass. The probability mass on the groundtruth is lower in this case since there is perceptual aliasing around the corners of the building that scan matching is unable to resolve completely. The maximum-likelihood sample with the distribution on landmark locations is shown in Figure 12. The metric map obtained from the Lu–Milios step is given in Figure 11(b).

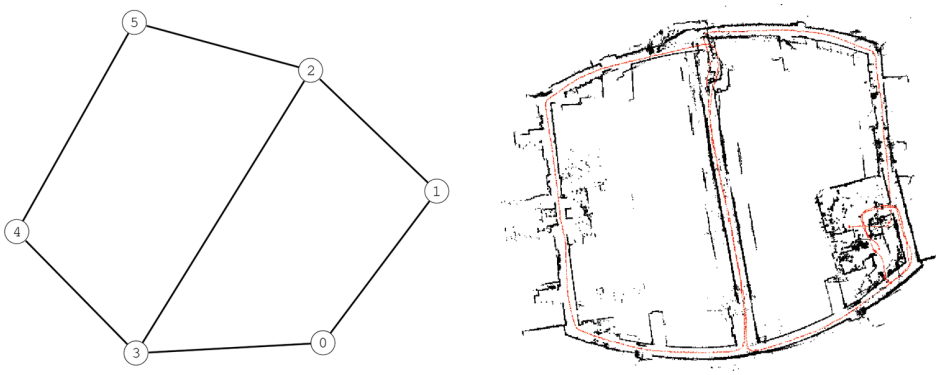
6.1. Landmarks at equidistant intervals

The perfect landmark detector assumed above is not available in practice. Hence, to validate the robustness of the OPTM algorithm we tested it with landmarks placed at equidistant intervals, which is the simplest and most redundant of landmark detection schemes, but has been used before in practical topological mapping schemes Tapus



(a) (b)

Fig. 10. (a) Schematic of the robot path overlaid on a floorplan of the TSRB environment. (b) Global metric map obtained using topological constraints and landmark locations for the second experiment. The robot path is the dotted (red) line.



(a) (b)

Fig. 11. (a) Groundtruth topology for third experiment on the CRB dataset. This receives 94% of the probability mass in the PTM. (b) Metric map obtained using topological constraints. The robot path is the dotted (red) line.

(2005). We now present experiments involving equidistant landmarks.

The first experiment was again in the TSRB dataset. A landmark was placed at every 10th image, i.e. approximately equally spaced in time. Hence, landmarks are closer spatially when the robot moves slowly and vice versa. A total of 35 landmarks were obtained using this scheme. The PTM at various stages of the incremental inference is illustrated in Figure 13. The final PTM contains only one topology, which is also the groundtruth. The covariances marking the uncertainty in landmark positions are also shown for the final topology.

The second experiment was performed using the well-known Intel dataset widely used in the SLAM literature (Hähnel et al. 2003b). The dataset consists of odometry and laser measurements, and the laser measurement model as described in Section 4.5 was used to obtain the result. Although appearance measurements are not available for this dataset, they can be left out of the computations simply by neglecting the appearance likelihood in the Bayes law equation (1). Landmarks were placed every 5 meters to obtain a total of 63 landmarks in the environment. The PTM contains nine topologies, and the most likely topology, which is also the groundtruth, obtains approximately

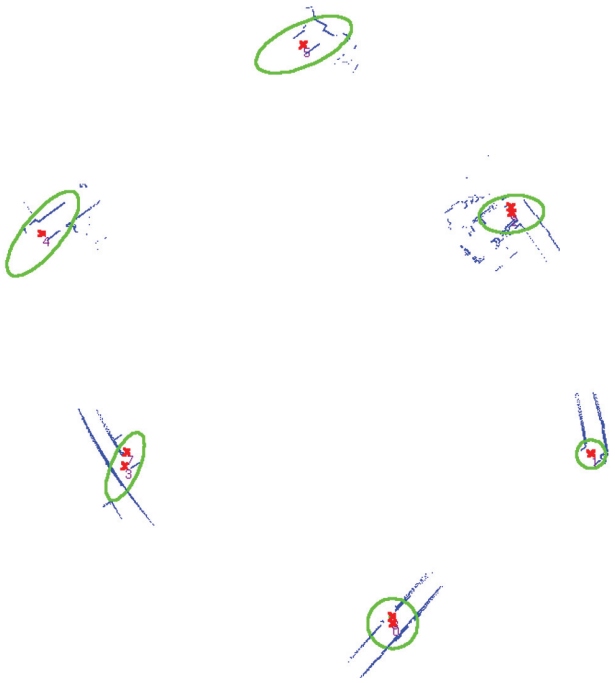


Fig. 12. MAP sample from the RBPF for the CRB experiment. Mean landmark locations are denoted as crosses while the ellipses denote marginal covariances.

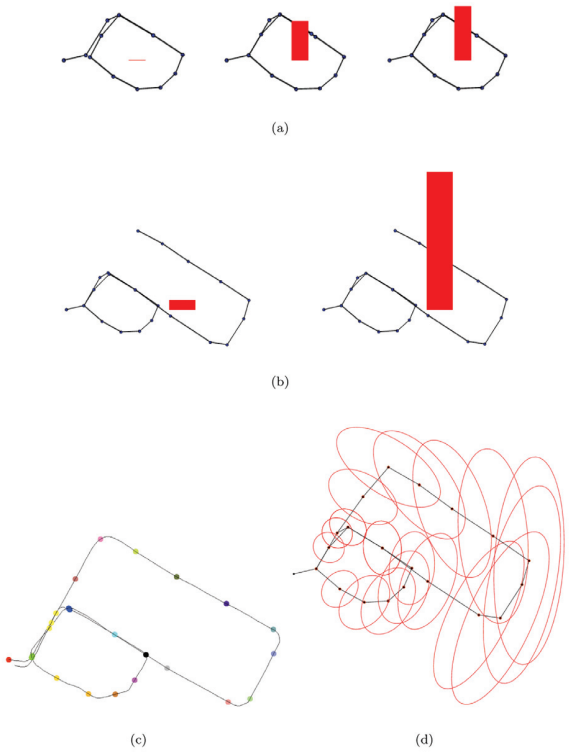


Fig. 13. PTM for the TSRB dataset with 35 landmarks placed equally in time. (a) PTM after 13 landmarks. (b) After 29 landmarks: he difference between the topologies is in the small triangular piece at the top of the smaller loop. (c) The final PTM after 35 landmarks contains only one topology, which is also the groundtruth. Landmarks are shown as small shaded (colored) circles with nodes corresponding to the same physical landmark shaded (colored) similarly. (d) 5σ covariance ellipses for the landmark locations of the groundtruth topology.

72% of the probability mass. The PTM along with the most likely topology is given in Figure 14. The metric map obtained from Hähnel et al. (2003a) is also shown for reference.

The smoothed trajectories for the above datasets were obtained using the optimization process described in Section 4.5 where it was used to compute the odometry likelihood, except that now all of the odometry measurements are used instead of just the compounded odometry between landmarks. The figure also illustrates the locations of the landmarks. Nodes classified as being the same physical landmark share the same shade (color).

6.2. Automatic Landmark Detection

We also evaluated the OPTM algorithm with the landmark detection algorithm described in Ranganathan and Dellaert (2009), which detects landmarks based on sudden changes in measurement values. We used the MIT Killian Court dataset (Bosse et al. 2004) which is another widely used dataset in the SLAM community. The dataset consists of 1,941 poses and corresponding laser scans. No appearance data is present in the dataset and the appearance likelihood was simply omitted in all of the OPTM calculations to obtain the results. The groundtruth metric map with laser scans and robot trajectory is shown in Figure 15 for reference. A total of 61 landmarks were detected using the landmark detection algorithm (Ranganathan and Dellaert 2009) and the PTM obtained using these landmarks, which also contains the groundtruth as the most likely topology, is shown in Figure 16. The groundtruth receives 81% of the probability mass. Figure 16(b) gives the trajectory smoothed with the topological constraints and also the color-coded nodes as before. Note that this trajectory does not look exactly the same as the groundtruth topology of Figure 16(a), especially at the lower right, because the landmark detector fails to detect a landmark in the segment connecting the small lower-right loop to the larger loop at the upper-right part of the image. Owing to this the two robot trails do not lie on top of each other in the groundtruth topology, which merely shows the landmark locations connected by edge. However, they do overlap in the smoothed trajectory map of Figure 16(b) due to the other neighboring topological constraints. Further, even though a few false positives are encountered in the landmark detection, the OPTM algorithm finds the groundtruth topology. This, and previous experiments using different landmark detection schemes, convincingly demonstrate that the OPTM algorithm is agnostic to the type of landmark detector used.

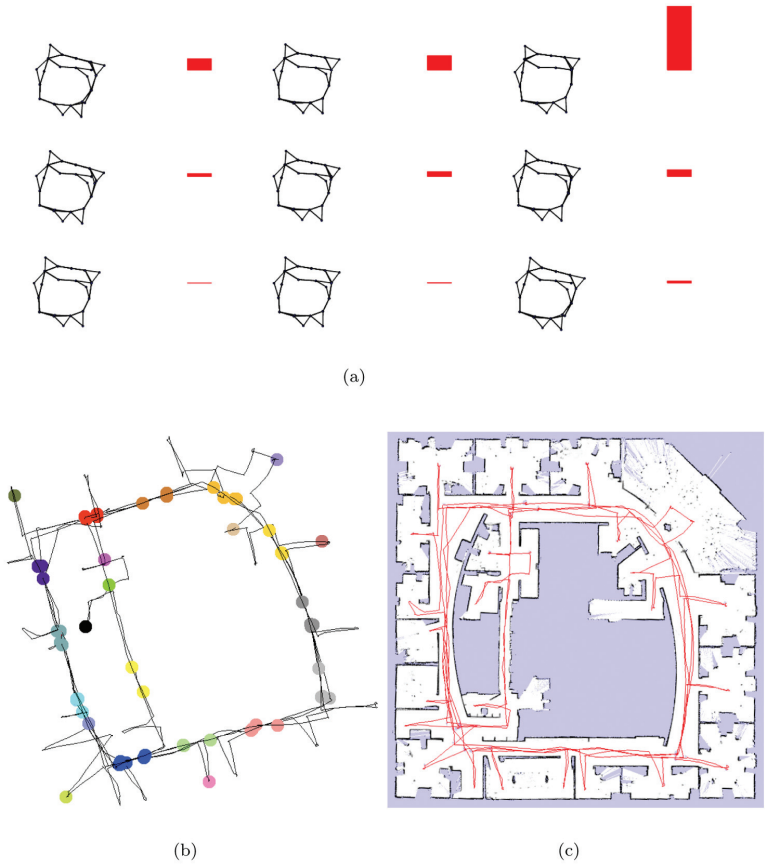


Fig. 14. (a) PTM (Bayesian posterior over loop closings) for the Intel dataset obtained using the OPTM algorithm. Sixty-three landmarks were placed in the environment at a distance of 5 meters from each other. The bars give the posterior probability of each topology. (b) Smoothed trajectory for the most likely topology, which is also the groundtruth topology and gets 72% of the probability mass. Landmarks are shown as shaded (colored) circles with nodes corresponding to the same physical landmark shaded (colored) similarly. (c) Metric map of the Intel lab given for reference.

Although the robot trajectory in this dataset spans an area of more than 200 m $\times$ 200 m and is considered challenging for metric mapping algorithms, it is a relatively easy sequence for performing topological mapping due to the wide separation between most landmarks, thus illustrating the advantage of a topological map over metric maps in this case.

6.3. *Timing results*

While the OPTM and the batch MCMC algorithm of Ranganathan et al. (2006) give similar mapping results, the OPTM is much faster. We present timing results for all of the datasets discussed above that confirm this statement. The fastest version of the PTM algorithm in Ranganathan et al. (2006) and the OPTM algorithm with the data-driven proposal were used to obtain the results given in Figure 17. Timing was obtained by running OCaml code on a 1.4 GHz dual-core machine with 2 GB of RAM. As can be seen, the OPTM is at least an order of magnitude faster than the batch MCMC algorithm, and considering the number of measurements in each of the datasets, is quite close to real-time in

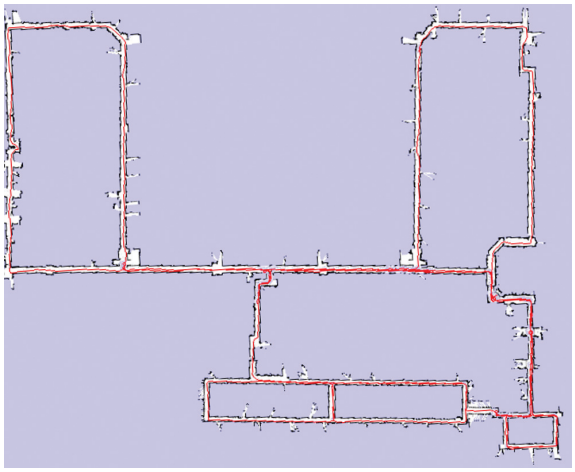
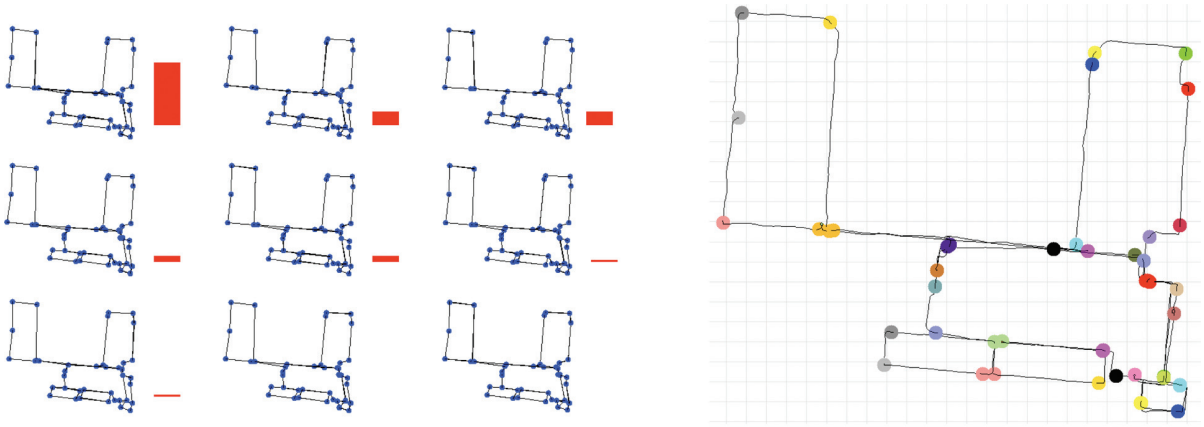


Fig. 15. Metric map of Killian Court dataset obtained from Bosse et al. (2004).



(a) (b)

Fig. 16. (a) PTM for the MIT Killian Court dataset with automatic landmark detection using Bayesian surprise. The topology at the top left with the maximum probability is the groundtruth. (b) The smoothed trajectory corresponding to the groundtruth topology.

	CRB	TSRB	Simulated	Killian Court	Intel Lab
No. of landmarks	9	12	33	61	63
Batch MCMC	9 s	9.8 s	56 s	292 s	409 s
OPTM	1.2 s	1.3 s	4.2 s	13.4 s	26.7 s

Fig. 17. Timing comparison of batch PTM and OPTM algorithms on various datasets.

operation. The Killian court dataset is mapped much faster than the Intel dataset, even though they have a similar number of landmarks, because the posterior contains fewer of topologies. This in turn is due to very little perceptual aliasing in the environment. Note that the inference procedure itself only has to be run when a landmark is detected, and not at every step.

6.4. Effect of the appearance model

The enhanced appearance model presented here is more resistant to perceptual aliasing, and hence provides better results compared with the simple Gaussian mixture over Fourier signatures used in Ranganathan et al. (2006). This is demonstrated by the more confident PTMs inferred by the OPTM with the bag-of-words appearance model. The output using the OPTM algorithm on the TSRB dataset with manually selected landmarks is a PTM where the groundtruth topology gets 98.5% of the probability mass. The corresponding probability for the RBPF using the Gaussian mixture appearance model is 95%. Similarly, the OPTM gives a probability of 94% for the groundtruth of the CRB dataset, while the probability using the simpler appearance model is 91%. Since the datasets used are exactly the same, we can attribute the better performance solely to the appearance model.

7. Conclusion and discussion

We have proposed a novel online particle filter-based loop-closing algorithm that is based on a systematic Bayesian framework for topological mapping. The framework works with various sensors and in diverse environments, as has been demonstrated in theory and in practice through experiments. While we have provided and tested sensor models for laser, odometry, and appearance, other sensors can be easily incorporated into the algorithm in a similar manner. Inference for loop closing with a large number of landmarks has also been demonstrated.

In future work we plane to incorporate more constraints such as planarity and limited types of junctions in man-made environments (Savelli and Kuipers 2004). Inclusion of these constraints would make the inference more efficient. The use of PTMs, in the form of a weighted ensemble of maps, for the purposes of planning and navigation, is also to be explored in the future.

A practical issue in the OPTM algorithm is the number of particles to use in inference. This is not only a question of efficiency but also one of correctness, since the algorithm may not converge to the true posterior with very few samples. In general, it is best to start the particle filtering algorithm with a reasonably large number of samples, say a hundred, and subsequently repeat the experiment with smaller numbers of samples to confirm convergence. In pathological cases, where perceptual aliasing is very high, even larger number of samples may be required.

The problem of convergence of the OPTM algorithm is a complicated one since convergence metrics that work in all cases are still an open topic of research in Monte Carlo sampling. However, a battery of convergence tests can be performed to ascertain convergence in the majority of cases. In the ideal scenario, every PTM algorithm should yield the same result since they are computing the same posterior over topologies. However, effects due to starting conditions and mixing speed are hard to avoid.

A converged sampler in the OPTM algorithm yields a theoretically correct posterior probability distribution, i.e. no algorithm using the same measurement models can obtain a better result. Hence, this can be used as a gold standard for comparison. In the absence of convergence, no such comparison of results in an objective manner is possible. A subjective test in such scenarios is to compare the topology with the highest probability to the groundtruth and ascertain whether they are the same.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest statement

None declared.

References

Angeli A, Filliat D, Doncieux S and Meyer J-A (2008) Fast and incremental method for loop-closure detection using bags of visual words. *IEEE Transactions on Robotics* 24:.

Aycard O, Charpillet F, Fohr D and Mari JF (1997) Place learning and recognition using hidden Markov models. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1741–1746.

Bailey T and Durrant-Whyte HF (2006) Simultaneous localisation and mapping (SLAM): Part II state of the art. *Robotics and Automation Magazine* (September).

Blackwell D and MacQueen JB (1973) Ferguson distributions via Polya urn schemes. *Annals of Statistics* 1: 353–355.

Blanco J-L, Fernandez-Madrigal J-A and Gonzalez J (2006) Consistent observation grouping for generating metric-topological maps that improves robot localization. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 818–823.

Blanco J-L, Fernandez-Madrigal J-A and Gonzalez J (2008) Towards a unified Bayesian approach to hybrid metric-topological SLAM. *IEEE Transactions on Robotics* 24: 259–270.

Bolles R and Fischler M (1981) A RANSAC-based approach to model fitting and its application to finding cylinders in range data. In *International Joint Conference on AI (IJCAI)*, Vancouver, BC, Canada, pp. 637–643.

Bosse MC, Newman PM, Leonard JJ and Teller S (2004) Simultaneous localization and map building in large-scale cyclic environments using the Atlas framework. *The International Journal of Robotics Research* 23: 1113–1139.

Burgard W, Fox D, Jans H, Matenar C and Thrun S (1999) Sonar-based mapping of large-scale mobile robot environments using EM. In *International Conference on Machine Learning (ICML)*, Bled, Slovenia, pp. 67–76.

Chen Y and Medioni G (1991) Object modelling by registration of multiple range images. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2724–2729.

Choset H and Nagatani K (2001) Topological simultaneous localization and mapping (SLAM): toward exact localization without explicit localization. *IEEE Transactions on Robotics and Automation* 17: 125–137.

Cummins M and Newman P (2008) FAB-MAP: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research* 27: 647–665.

Dedeoglu G, Mataric M and Sukhatme H (1999) Incremental, online topological map building with a mobile robot. In *Proceedings of Mobile Robots*.

Dellaert F (2001) *Monte Carlo EM for Data Association and its Applications in Computer Vision*. PhD thesis, School of Computer Science, Carnegie Mellon, September 2001. Also available as Technical Report CMU-CS-01-153.

Dellaert F, Seitz SM, Thorpe CE and Thrun S (2003) EM, MCMC, and chain flipping for structure from motion with unknown correspondence. *Machine Learning* 50: 45–71.

Doucet A, Godsill S and Andrieu C (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* 10: 197–208.

Dudek G, Hadjres S and Freedman P (1993) Using local information in a non-local way for mapping graph-like worlds. In *International Joint Conference on AI (IJCAI)*, pp. 1639–1645.

Dudek G and Jugessur D (2000) Robust place recognition using local appearance based methods. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1030–1035.

Eliazar A and Parr R (2003) DP-SLAM: Fast, robust simultaneous localization and mapping without predetermined landmarks. In *International Joint Conference on AI (IJCAI)*.

Fearnhead P and Clifford P (2003) Online inference for hidden Markov models. *Journal of the Royal Statistical Society: Series B*, 65: 887–899.

Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. *Annals of Statistics* 1: 209–230.

Goedemé T, Nuttin M, Tuytelaars T and Van Gool L (2007) Omnidirectional vision based topological navigation. *International Journal of Computer Vision* 74: 219–236.

Gutierrez-Osuna R and Luo RC (1996) Lola: Probabilistic navigation for topological maps. *AI Magazine* 17: 55–62.

Gutmann J-S and Konolige K (2000) Incremental mapping of large cyclic environments. In *IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA)*, pp. 318–325.

Hähnel D, Burgard W, Fox D and Thrun S (2003a) A highly efficient FastSLAM algorithm for generating cyclic maps of large-scale environments from raw laser range measurements. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 206–211.

Hähnel D, Burgard W, Wegbreit B and Thrun S (2003b) Towards lazy data association in SLAM. In *Proceedings of the 11th International Symposium of Robotics Research (ISRR'03)*, Sienna, Italy. Berlin: Springer.

Kaelbling L, Cassandra AR and Kurien JA (1996) Acting under uncertainty: Discrete Bayesian models for mobile-robot navigation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Kaess M and Dellaert F (2005) A Markov chain Monte Carlo approach to closing the loop in SLAM. In *IEEE International Conference on Robotics and Automation (ICRA)*, Barcelona, Spain, April 2005, pp. 645–650.

Kuipers B and Beeson P (2002) Bootstrap learning for place recognition. In *Proceedings 19th AAAI National Conference on AI*, Edmonton, Alberta, Canada, 2002, pp. 174–180.

Kuipers BJ and Byun Y-T (1991) A robot exploration and mapping strategy based on a semantic hierarchy of spatial