ELSEVIER

Artificial Intelligence 128 (2001) 99–141

Artificial Intelligence

www.elsevier.com/locate/artint

Robust Monte Carlo localization for mobile robots

Sebastian Thrun [a,*], Dieter Fox [b], Wolfram Burgard [c], Frank Dellaert [a]

[a] *School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[b] *Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195, USA*
[c] *Computer Science Department, University of Freiburg, Freiburg, Germany*

Received 20 April 2000

Abstract

Mobile robot localization is the problem of determining a robot's pose from sensor data. This article presents a family of probabilistic localization algorithms known as Monte Carlo Localization (MCL). MCL algorithms represent a robot's belief by a set of weighted hypotheses (samples), which approximate the posterior under a common Bayesian formulation of the localization problem. Building on the basic MCL algorithm, this article develops a more robust algorithm called Mixture-MCL, which integrates two complimentary ways of generating samples in the estimation. To apply this algorithm to mobile robots equipped with range finders, a kernel density tree is learned that permits fast sampling. Systematic empirical results illustrate the robustness and computational efficiency of the approach. © 2001 Published by Elsevier Science B.V.

Keywords: Mobile robots; Localization; Position estimation; Particle filters; Kernel density trees

1. Introduction

Mobile robot localization is the problem of estimating a robot's pose (location, orientation) relative to its environment. The localization problem is a key problem in mobile robotics. It plays a pivotal role in various successful mobile robot systems (see e.g., [10,25,31,45,59,65,74] and various chapters in [4,41]). Occasionally, it has been referred to as "the most fundamental problem to providing a mobile robot with autonomous capabilities" [8].

The mobile robot localization problem comes in many different flavors [4,24]. The most simple localization problem—which has received by far the most attention in the literature—is *position tracking* [4,64,74,75]. Here the initial robot pose is known, and

* Corresponding author.
E-mail address: thrun@cs.cmu.edu (S. Thrun).

0004-3702/01/$ – see front matter
PII: S0004-3702(01)00069-8

the problem is to compensate incremental errors in a robot's odometry. Algorithms for position tracking often make restrictive assumptions on the size of the error and the shape of the robot's uncertainty, required by a range of existing localization algorithms. More challenging is the *global localization problem* [6,34,61], where a robot is not told its initial pose but instead has to determine it from scratch. The global localization problem is more difficult, since the error in the robot's estimate cannot be assumed to be small. Consequently, a robot should be able to handle multiple, distinct hypotheses. Even more difficult is the *kidnapped robot problem* [20,24], in which a well-localized robot is teleported to some other place without being told. This problem differs from the global localization problem in that the robot might firmly believe itself to be somewhere else at the time of the kidnapping. The kidnapped robot problem is often used to test a robot's ability to recover from catastrophic localization failures. Finally, all these problems are particularly hard in dynamic environments, e.g., if robots operate in the proximity of people who corrupt the robot's sensor measurements [5,71].

The vast majority of existing algorithms address only the position tracking problem (see e.g., the review [4]). The nature of small, incremental errors makes algorithms such as Kalman filters [28,37,47,68] applicable, which have been successfully applied in a range of fielded systems (e.g., [27,42,44,63]). Kalman filters estimate posterior distributions of robot poses conditioned on sensor data. Exploiting a range of restrictive assumptions—such as Gaussian noise and Gaussian-distributed initial uncertainty—they represent posteriors by Gaussians. Kalman filters offer an elegant and efficient algorithm for localization. However, the restrictive nature of the belief representation makes plain Kalman filters inapplicable to global localization problems.

This limitation is overcome by two related families of algorithms: localization with multi-hypothesis Kalman filters and *Markov localization*. Multi-hypothesis Kalman filters represent beliefs using mixtures of Gaussians [9,34,60,61], thereby enabling them to pursue multiple, distinct hypotheses, each of which is represented by a separate Gaussian. However, this approach inherits from Kalman filters the Gaussian noise assumption. To meet this assumption, virtually all practical implementations extract low-dimensional features from the sensor data, thereby ignoring much of the information acquired by a robot's sensors. Markov localization algorithms, in contrast, represent beliefs by piecewise constant functions (histograms) over the space of all possible poses [7,24,30,36,40,50, 55,56,66,70]. Just like Gaussian mixtures, piecewise constant functions are capable of representing complex, multi-modal representations. Some of these algorithms also rely on features [36,40,50,55,66,70], hence are subject to similar shortcomings as the algorithms based on multi-hypothesis Kalman filters. Others localize robots based on raw sensor data with non-Gaussian noise distributions [7,24]. However, accommodating raw sensor data requires fine-grained representations, which impose significant computational burdens. To overcome this limitation, researchers have proposed selective updating algorithms [24] and tree-based representations that dynamically change their resolution [6]. It is remarkable that all of these algorithms share the same probabilistic basis. They all estimate posterior distributions over poses under certain independence assumptions—which will also be the case for the approach presented in this article.

This article presents a probabilistic localization algorithm called Monte Carlo localization (MCL) [13,21]. MCL solves the global localization and kidnapped robot problem

in a robust and efficient way. It can accommodate arbitrary noise distributions (and non-linearities in robot motion and perception). Thus, MCL avoids a need to extract features from the sensor data.

The key idea of MCL is to represent the belief by a set of samples (also called: *particles*), drawn according to the posterior distribution over robot poses. In other words, rather than approximating posteriors in parametric form, as is the case for Kalman filter and Markov localization algorithms, MCL represents the posteriors by a random collection of weighted particles which approximates the desired distribution [62]. The idea of estimating state recursively using particles is not new, although most work on this topic is very recent. In the statistical literature, it is known as *particle filters* [17, 18,46,58], and recently computer vision researchers have proposed the same algorithm under the name of *condensation algorithm* [33]. Within the context of localization, the particle representation has a range of characteristics that sets it aside from previous approaches:

(1) Particle filters can accommodate (almost) arbitrary sensor characteristics, motion dynamics, and noise distributions.
(2) Particle filters are universal density approximators, weakening the restrictive assumptions on the shape of the posterior density when compared to previous parametric approaches.
(3) Particle filters focus computational resources in areas that are most relevant, by sampling in proportion to the posterior likelihood.
(4) By controlling the number of samples on-line, particle filters can adapt to the available computational resources. The same code can, thus, be executed on computers with vastly different speed without modification.
(5) Finally, participle filters are surprisingly easy to implement, which makes them an attractive paradigm for mobile robot localization. Consequently, MCL has already been adopted by several research teams [16,43], who have extended the basic paradigm in interesting new ways.

However, there are pitfalls, too, arising from the stochastic nature of the approximation. Some of these pitfalls are obvious: For example, if the sample set size is small, a well-localized robot might lose track of its position just because MCL fails to generate a sample in the right location. The regular MCL algorithm is also unfit for the kidnapped robot problem, since there might be no surviving samples nearby the robot's new pose after it has been kidnapped. Somewhat counter-intuitive is the fact that the basic algorithm degrades poorly when sensors are too accurate. In the extreme, regular MCL will fail with perfect, noise-free sensors. All these problems can be overcome, e.g., by augmenting the sample set through uniformly distributed samples [21], generating samples consistent with the most recent sensor reading [43] (an idea familiar from multi-hypothesis Kalman filtering [1,34, 61]), or assuming a higher level of sensor noise than actually is the case. While these extensions yield improved performance, they are mathematically questionable. In particular, these extensions do not approximate the correct density; which makes the interpretation of their results difficult.

To overcome these problems, this article describes an extension of MCL closely related to [43], called *Mixture-MCL* [72]. Mixture-MCL addresses all these problems in a way that is mathematically motivated. The key idea is to modify the way samples are generated

in MCL. Mixture-MCL combines regular MCL sampling with a "dual" of MCL, which basically inverts MCL's sampling process. More specifically, while regular MCL first guesses a new pose using odometry, then uses the sensor measurements to assess the "importance" of this sample, dual MCL guesses poses using the most recent sensor measurement, then uses odometry to assess the compliance of this guess with the robot's previous belief and odometry data. Neither of these sampling methodologies alone is sufficient; in combination, however, they work very well. In particular, Mixture-MCL works well if the sample set size is small (e.g., 50 samples), it recovers faster from robot kidnapping than any previous variation of MCL, and it also works well when sensor models are too narrow for regular MCL. Thus, from a performance point of view, Mixture-MCL is uniformly superior to regular MCL and particle filters.

The key *disadvantage* of Mixture-MCL is a requirement for a sensor model that permits fast sampling of poses. While in certain cases, such a model can trivially be obtained, in others, such as the navigation domains studied here and in [24], it cannot. To overcome this difficulty, our approach uses sufficient statistics and density trees to learn a sampling model from data. More specifically, in a pre-processing phase sensor readings are mapped into a set of discriminating features, and potential robot poses are then drawn randomly using trees generated. Once the tree has been constructed, dual sampling can be done very efficiently.

To shed light onto the performance of Mixture-MCL in practice, empirical results are presented using a robot simulator and data collected by physical robots. Simulation is used since it allows us to systematically vary key parameters such as the sensor noise, thereby enabling us to characterize the degradation of MCL in extreme situations. To verify the experimental findings obtained with simulation, Mixture-MCL is also applied to two extensive data sets gathered in a public museum (a Smithsonian museum in Washington, DC), where during a two-week period in the fall of 1998 our mobile robot Minerva gave interactive tours to thousands of visitors [71]. One of the data set comprises laser range data, where a metric map of the museum is used for localization [71]. The other data set contains image segments recorded with a camera pointed towards the museum's ceiling, using a large-scale ceiling mosaic for cross-referencing the robot's position [14]. In the past, these data have been used as benchmark, since localization in this crowded and feature-impoverished museum is a challenging problem. Our experiments suggest that our new MCL algorithm is highly efficient and accurate.

The remainder of this article is organized as follows. Section 2 introduces the regular MCL algorithm, which includes a mathematical derivation from first principles and an experimental characterization of MCL in practice. The section also compares MCL with grid-based Markov localization [24], an alternative localization algorithms capable of global localization. Section 3 presents examples where regular MCL performs poorly, along with a brief analysis of the underlying causes. This section is followed by the description of dual MCL and Mixture-MCL in Section 4. Section 5 describes our approach to learning trees for efficient sampling in dual MCL. Experimental results are given in Section 6. Finally, we conclude this article by a description of related work in Section 7, and a discussion of the strengths and weaknesses of Mixture-MCL (Section 8).

2. Monte Carlo localization

2.1. Bayes filtering

MCL is a recursive Bayes filter that estimates the posterior distribution of robot poses conditioned on sensor data. Bayes filters address the problem of estimating the state x of a dynamical system (partially observable Markov chain) from sensor measurements. For example, in mobile robot localization the dynamical system is a mobile robot and its environment, the state is the robot's pose therein (often specified by a position in a two-dimensional Cartesian space and the robot's heading direction), and measurements may include range measurements, camera images, and odometry readings. Bayes filters assume that the environment is *Markov*, that is, past and future data are (conditionally) independent if one knows the current state. The Markov assumption will be made more explicit below.

The key idea of Bayes filtering is to estimate a probability density over the state space conditioned on the data. This posterior is typically called the *belief* and is denoted

$$Bel(x_t) = p(x_t \mid d_{0\ldots t}).$$

Here x denotes the state, x_t is the state at time t, and $d_{0\ldots t}$ denotes the data starting at time 0 up to time t. For mobile robots, we distinguish two types of data: *perceptual data* such as laser range measurements, and *odometry data*, which carry information about robot motion. Denoting the former by o (for *observation*) and the latter by a (for *action*), we have

$$Bel(x_t) = p(x_t \mid o_t, a_{t-1}, o_{t-1}, a_{t-2}, \ldots, o_0). \tag{1}$$

Without loss of generality, we assume that observations and actions arrive in an alternating sequence. Notice that we will use a_{t-1} to refer to the odometry reading that measures the motion that occurred in the time interval $[t-1; t]$, to illustrate that the motion is the result of the control action asserted at time $t-1$.

Bayes filters estimate the belief *recursively*. The *initial* belief characterizes the *initial* knowledge about the system state. In the absence of such knowledge, it is typically initialized by a *uniform distribution* over the state space. In mobile robot localization, a uniform initial distribution corresponds to the global localization problem, where the initial robot pose is unknown.

To derive a recursive update equation, we observe that expression (1) can be transformed by Bayes rule to

$$Bel(x_t) = \frac{p(o_t \mid x_t, a_{t-1}, \ldots, o_0)\, p(x_t \mid a_{t-1}, \ldots, o_0)}{p(o_t \mid a_{t-1}, \ldots, o_0)}. \tag{2}$$

Because the denominator is a constant relative to the variable x_t, Bayes rule is usually written as

$$Bel(x_t) = \eta p(o_t \mid x_t, a_{t-1}, \ldots, o_0)\, p(x_t \mid a_{t-1}, \ldots, o_0), \tag{3}$$

where η is the normalization constant

$$\eta = p(o_t \mid a_{t-1}, \ldots, o_0)^{-1}. \tag{4}$$

As noticed above, Bayes filters rest on the assumption that future data is independent of past data given knowledge of the current state—an assumption typically referred to as the *Markov assumption*. Put mathematically, the Markov assumption implies

$$p(o_t \mid x_t, a_{t-1}, \ldots, o_0) = p(o_t \mid x_t) \tag{5}$$

and hence our target expression (3) can be simplified to:

$$Bel(x_t) = \eta p(o_t \mid x_t)\, p(x_t \mid a_{t-1}, \ldots, o_0).$$

We will now expand the rightmost term by integrating over the state at time $t-1$:

$$Bel(x_t) = \eta p(o_t \mid x_t) \int p(x_t \mid x_{t-1}, a_{t-1}, \ldots, o_0)\, p(x_{t-1} \mid a_{t-1}, \ldots, o_0)\, \mathrm{d}x_{t-1}. \tag{6}$$

Again, we can exploit the Markov assumption to simplify $p(x_t \mid x_{t-1}, a_{t-1}, \ldots, o_0)$:

$$p(x_t \mid x_{t-1}, a_{t-1}, \ldots, o_0) = p(x_t \mid x_{t-1}, a_{t-1}) \tag{7}$$

which gives us the following expression:

$$Bel(x_t) = \eta p(o_t \mid x_t) \int p(x_t \mid x_{t-1}, a_{t-1})\, p(x_{t-1} \mid a_{t-1}, \ldots, o_0)\, \mathrm{d}x_{t-1}. \tag{8}$$

Substituting the basic definition of the belief *Bel* back into (8), we obtain the important recursive equation

$$Bel(x_t) = \eta p(o_t \mid x_t) \int p(x_t \mid x_{t-1}, a_{t-1}) Bel(x_{t-1})\, \mathrm{d}x_{t-1}. \tag{9}$$

This equation is the recursive update equation in Bayes filters. Together with the initial belief, it defines a recursive estimator for the state of a partially observable system. This equation is of central importance in this article, as it is the basis for various MCL algorithms studied here.

To implement (9), one needs to know two conditional densities: the probability $p(x_t \mid x_{t-1}, a_{t-1})$, which we will refer to as *next state density* or simply *motion model*, and the density $p(o_t \mid x_t)$, which we will call *perceptual model* or *sensor model*. Both models are typically *stationary* (also called: *time-invariant*), that is, they do not depend on the specific time t. This stationarity allows us to simplify the notation by denoting these models $p(x' \mid x, a)$, and $p(o \mid x)$, respectively.

2.2. Probabilistic models for localization

The nature of the models $p(x' \mid x, a)$ and $p(o \mid x)$ depends on the specific estimation problem. In mobile robot localization, which is the focus of this article, both models are relatively straightforward and can be implemented in a few lines of code. The specific probabilistic models used in our implementation have been described in depth elsewhere [24]; therefore, we will only provide an informal account.

The motion model, $p(x' \mid x, a)$, is a probabilistic generalization of robot kinematics [10,73]. As noticed above, for a robot operating in the plane the poses x and x' are

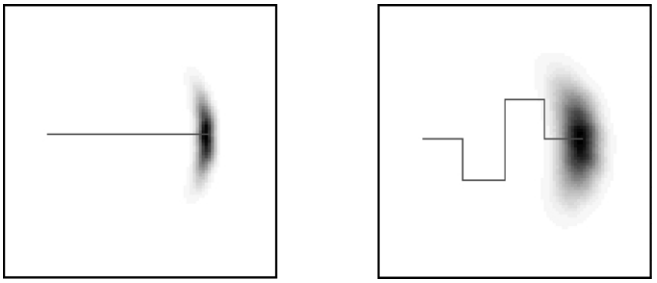


Fig. 1. The density $p(x' \mid x, a)$ after moving 40 meter (left diagram) and 80 meter (right diagram). The darker a pose, the more likely it is.

three-dimensional variables. Each pose comprises a robot's two-dimensional Cartesian coordinates and its heading direction (orientation, bearing). The value of a may be an odometry reading or a control command, both of which characterize the change of pose. In robotics, change of pose is called *kinematics*. The conventional kinematic equations, however, describe only the *expected* pose x' that an ideal, noise-free robot would attain starting at x, and after moving as specified by a. Of course, physical robot motion is erroneous; thus, the pose x' is uncertain. To account for this inherent uncertainty, the probabilistic motion model $p(x' \mid x, a)$ describes a posterior density over possible successors x'. Noise is typically modeled by zero centered, Gaussian noise that is added to the translation and rotation components in the odometry measurements [24]. Thus, $p(x' \mid x, a)$ generalizes exact mobile robot kinematics typically described in robot textbooks [10,73] by a probabilistic component.

Fig. 1 shows two examples of $p(x' \mid x, a)$. In both examples, the initial pose x is shown on the left, and the solid line depicts the odometry data as measured by the robot. The grayly shaded area on the right depicts the density $p(x' \mid x, a)$: the darker a pose, the more likely it is. A comparison of both diagrams reveals that the margin of uncertainty depends on the overall motion: Even though the pose of a noise-free robot are the same for both motion segments, the uncertainty in the right diagram is larger due to the longer overall distance traversed by the robot.

For the MCL algorithm described further below, one does not need a closed-form description of the motion model $p(x' \mid x, a)$. Instead, a *sampling model* of $p(x' \mid x, a)$ suffices. A sampling model is a routine that accepts x and a as an input and generates random poses x' distributed according to $p(x' \mid x, a)$. Sampling models are usually easier to code than routines that compute densities in closed form. Fig. 2 shows a sample model of $p(x' \mid x, a)$, applied to a sequence of odometry measurements, as indicated by the solid line. As is easy to be seen, the sequence of particle sets approximates the densities of a robot that only measures odometry.

Let us now turn our attention to the perceptual model, $p(o \mid x)$. Mobile robots are commonly equipped with range finders, such as ultrasonic transducers (sonar sensors) or laser range finders. Fig. 3(a) shows an example of a laser range scan, obtained with an RWI B21 robot in an environment whose approximate shape is also shown in Fig. 3(a). Notice that the range finder emits a plateau of light that covers a horizontal 180 degree range, for which it measures the distance to the nearest objects.

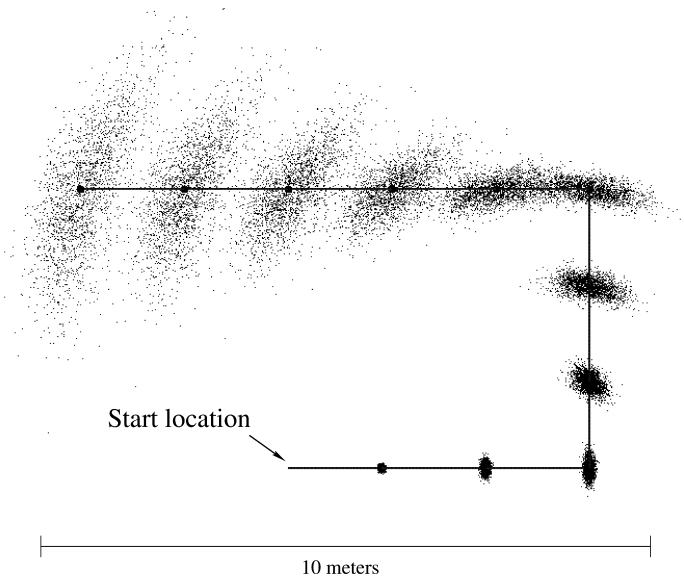


Fig. 2. Sampling-based approximation of the position belief for a robot that only measures odometry. The solid line displays the actions, and the samples represent the robot's belief at different points in time.

For range finders, we decompose the problem of computing $p(o \mid x)$ into three parts:

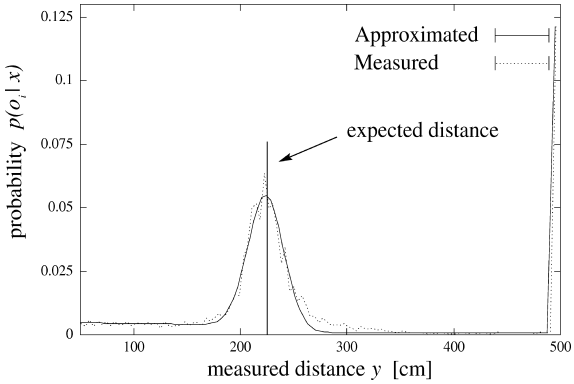
(1) the computation of the value a noise-free sensor would generate;
(2) the modeling of sensor noise; and
(3) the integration of many individual sensor beams into a single density value.

Assume the robot's pose is x, and let o_i denote an individual sensor beam with bearing α_i relative to the robot. Let $g(x, \alpha_i)$ denote the measurement of an ideal, noise-free sensor whose relative bearing is α_i. Since we assume that the robot is given a map of the environment such as the one shown in Fig. 3(a), $g(x, \alpha_i)$ can be computed using *ray tracing* [49]. We assume that this "expected" distance $g(x, \alpha_i)$ is a sufficient statistic for the probability $p(o_i \mid x)$, that is

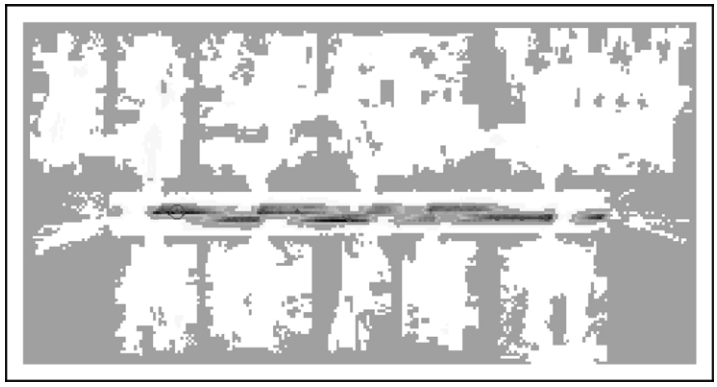
$$p(o_i \mid x) = p\big(o_i \mid g(x, \alpha_i)\big). \tag{10}$$

The exact density $p(o_i \mid x)$ is shown in Fig. 3(b). This density is a mixture of three densities: a Gaussian centered at $g(x, \alpha_i)$ that models the event of measuring the correct distance with small added Gaussian noise, an exponential density that models random readings as often caused by people, and a discrete large probability (mathematically modeled by a narrow uniform density) that models max-range measurements, which frequently occur when a range sensor fails to detect an object. The specific parameters of the density in Fig. 3(b) have been estimated using an algorithm similar to EM [15,52], which starts with a crude initial model and iteratively labels several million measurements collected in the Smithsonian museum, while refining the model. A smoothed version of these data is also shown in Fig. 3(b), illustrating that our probabilistic model is highly accurate.

(a) laser scan and map



(b) sensor model $p(o_i \mid x)$



(c) probability distribution for different poses

Fig. 3. (a) Laser range scan, projected into a map. (b) The density $p(o \mid x)$, where the peak corresponds to the distance an ideal, noise-free sensor would measure. (c) $p(o \mid x)$ for the scan shown in (a). Based on a single sensor scan, the robot assigns high likelihood for being somewhere in the main corridor.

Finally, the individual density values $p(o_i \mid x)$ are integrated multiplicatively, assuming conditional independence between the individual measurements:

$$p(o \mid x) = \prod_i p(o_i \mid x). \tag{11}$$

Clearly, this conditional independence can be violated in the presence of people (which often block more than one sensor beam). In such cases, it might be advisable to subsample the sensor readings and use a reduced set for localization [24].

Fig. 3(c) depicts $p(o \mid x)$ for the sensor scan shown in Fig. 3(a) and the map shown in gray in Fig. 3(c). Most of the probability mass is found in the corridor region, which reflects the fact that the specific sensor measurement is more likely to have originated in the corridor than in one of the rooms. The probability mass roughly lies on two bands, and otherwise is spread through most of the corridor. The density shown in Fig. 3(c) is equivalent to the posterior belief of a globally localizing robot after perceiving only one sensor measurement.

2.3. Particle approximation

If the state space is continuous, as is the case in mobile robot localization, implementing (9) is *not* a trivial matter—particularly if one is concerned about efficiency.

The idea of MCL (and other particle filter algorithms) is to represent the belief $Bel(x)$ by a set of m weighted samples distributed according to $Bel(x)$:

$$Bel(x) \approx \left\{x^{(i)}, w^{(i)}\right\}_{i=1,\ldots,m}.$$

Here each $x^{(i)}$ is a *sample* of the random variable x, that is, a hypothesized state (pose). The non-negative numerical parameters $w^{(i)}$ are called *importance factors*, which throughout this paper will be such that they sum to 1 (although this is not strictly required in particle filtering). As the name suggests, the importance factors determine the weight (=importance) of each sample [62]. The set of samples, thus, define a discrete probability function that approximates the continuous belief $Bel(x)$.

The initial set of samples represents the initial knowledge $Bel(x_0)$ about the state of the dynamical system. For example, in global mobile robot localization, the initial belief is a set of poses drawn according to a uniform distribution over the robot's universe, annotated by the uniform importance factor $1/m$. If the initial pose is known up to some small margin of error, $Bel(x_0)$ may be initialized by samples drawn from a narrow Gaussian centered on the correct pose.

The recursive update is realized in three steps.

(1) Sample $x_{t-1}^{(i)} \sim Bel(x_{t-1})$ from the (weighted) sample set representing $Bel(x_{t-1})$. Each such particle $x_{t-1}^{(i)}$ is distributed according to the belief distribution $Bel(x_{t-1})$.

(2) Sample $x_t^{(i)} \sim p(x_t \mid x_{t-1}^{(i)}, a_{t-1})$. Obviously, the pair $\langle x_t^{(i)}, x_{t-1}^{(i)} \rangle$ is distributed according to the product distribution

$$q_t := p(x_t \mid x_{t-1}, a_{t-1}) \times Bel(x_{t-1}). \tag{12}$$

In accordance with the literature on the SIR algorithm (short for: Sampling importance resampling) [62,67,69], we will refer to this distribution q_t as the

proposal distribution. Its role is to "propose" samples of the desired posterior distribution, which is given in Eq. (9); however, it is not equivalent to the desired posterior.

(3) Finally, correct for the mismatch between the proposal distribution q_t and the desired distribution specified in Eq. (9) and restated here for the sample pair $\langle x_{t-1}^{(i)}.x_t^{(i)}\rangle$:

$$\eta p\big(o_t \mid x_t^{(i)}\big)\, p\big(x_t^{(i)} \mid x_{t-1}^{(i)}, a_{t-1}\big) Bel\big(x_{t-1}^{(i)}\big). \tag{13}$$

This mismatch is accounted for by the following (non-normalized) importance factor, which is assigned to the ith sample:

$$w^{(i)} = p\big(o_t \mid x_t^{(i)}\big). \tag{14}$$

Following [62], the importance factor is obtained as the quotient of the target distribution (13) and the proposal distribution (12), up to a constant scaling factor:

$$\frac{\eta p(o_t \mid x_t^{(i)})\, p(x_t^{(i)} \mid x_{t-1}^{(i)}, a_{t-1}) Bel(x_{t-1}^{(i)})}{p(x_t^{(i)} \mid x_{t-1}^{(i)}, a_{t-1}) Bel(x_{t-1}^{(i)})} = \eta p\big(o_t \mid x_t^{(i)}\big). \tag{15}$$

Notice that this quotient is proportional to $w^{(i)}$, since η is a constant.

The sampling routine is repeated m times, producing a set of m weighted samples $x_t^{(i)}$ (with $i = 1, \ldots, m$). Afterwards, the importance factors are normalized so that they sum up to 1 and hence define a discrete probability distribution.

Table 1 summarizes the MCL algorithm. It is known [69] that under mild assumptions (which hold in our work), the sample set converges to the true posterior $Bel(x_t)$ as m goes to infinity, with a convergence speed in $O(1/\sqrt{m})$. The speed may vary by a constant factor, which can vary drastically depending on the proposal distribution. Due to the normalization, the particle filter is only asymptotically unbiased. Care has to be taken if the number of samples is extremely small (e.g., less than 10), as the bias increases as

Table 1
The MCL algorithm

Algorithm MCL(X, a, o)
$X' = \emptyset$
for $i = 0$ to m do
generate random x from X according to $w_1, \ldots, w_m$
generate random $x' \sim p(x' \mid a, x)$
$w' = p(o \mid x')$
add $\langle x', w'\rangle$ to X'
endfor
normalize the importance factors w' in X'
return X'

the sample set size decreases. In the extreme case of $m = 1$, the measurements o_t will plainly be ignored, and the resulting expectation of this single sample is heavily biased by the prior. In all our implementations, however, the number of samples is sufficiently large.

2.4. Examples

We performed systematic experiments in a range of different settings to evaluate the performance of MCL in practice.

2.5. Simulation

Simulation was employed for evaluation since it allows us to freely vary key parameters, such as the amount of sensor noise. Further below, we will make use of this freedom to characterize situations in which MCL performs poorly.

Fig. 4 shows an example in which a simulated mobile robot localizes an object in 3D. In our simulation, this robot can detect the (approximate) location of the object in the image taken by its camera, but the lack of depth estimation in mono-vision makes it impossible to localize the object from a single camera image. Instead, the simulated robot has to view the object from multiple viewpoints. However, changing the viewpoint introduces additional uncertainty, as robot motion is inaccurate. Additionally, in our simulation, the visual field of the robot is limited to a narrow region in front of the robot, which further

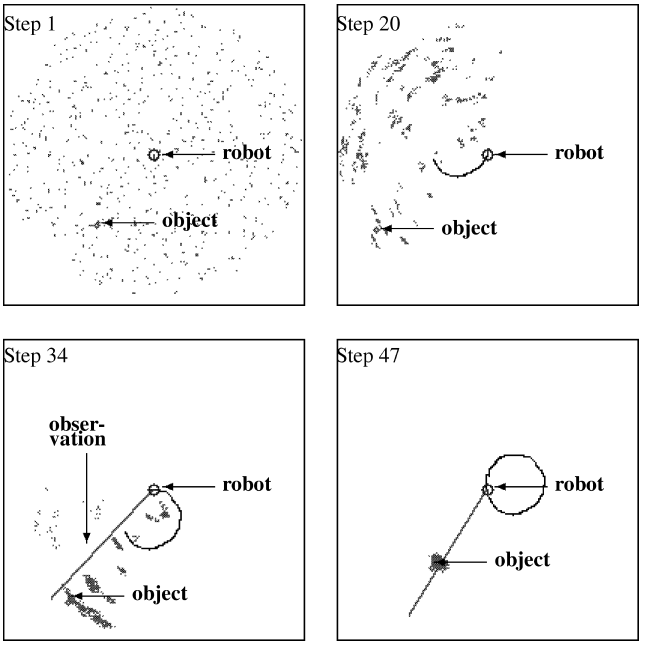


Fig. 4. Successful localization sequence for our robot simulation of object localization. The final error is 36.0 cm.

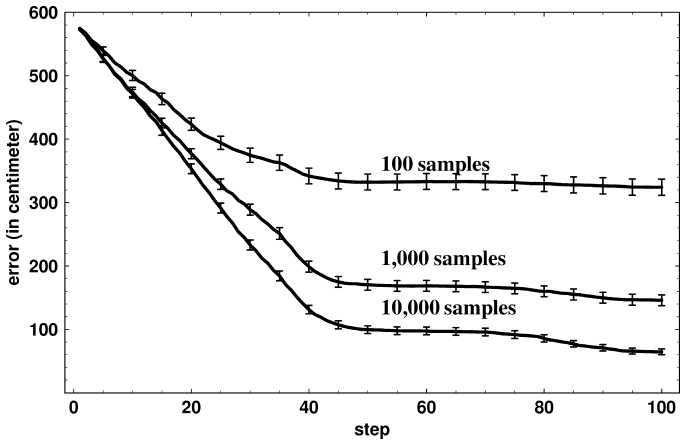


Fig. 5. Average error of MCL as a function of the number of simulated robot steps/measurements.

complicates the object localization problem. Our noise simulation includes a simulation of measurement noise, false positive measurements (phantom detections) and false negative measurements (failures to detect the target object). To enable us to systematically vary key parameters such as the perceptual noise, our results use a mobile robot simulator that models control error (reminiscent of an RWI B21 robot) and noise in perception (Gaussian position noise, false negatives, and false positives). Notice that this task is more difficult than the first one, due to the impoverished nature of the robot sensors and the large number of symmetries.

Fig. 4 depicts different states of global object localization. Initially, the pose of the object is unknown, as represented by the uniform sample set in Fig. 4 (first diagram). As the simulated robot turns in a wide circle, unable to see the object (despite a handful of phantom detections), the samples gradually populate the unexplored part of the state space (Fig. 4, second diagram). The third diagram shows the first "correct" object sighting, which reinforces the samples that happen to be close to the correct object pose. A few measurements later, repetitive sightings of the object lead to a posterior shown in the fourth diagram of Fig. 4. Fig. 5 shows systematic error curves for MCL in global localization for different sample set sizes m, averaged over 1,000 individual experiments. The bars in this figure are confidence intervals at the 95% level.

The reader should notice that these results have been obtained for a perceptual noise level of 20% (for both false-negative and false-positive) and an additional position noise that is Gaussian-distributed with a variance of 10 degrees. For our existing vision system, these error rates are much lower, but may not necessarily be independent.

2.6. Robot with sonar sensors

Fig. 6 shows an example of MCL in the context of localizing a mobile robot globally in an office environment. This robot, called Minerva, is an RWI B18 robot equipped with sonar range finders. It is given a map of the environment. In Fig. 6(a), the robot is globally

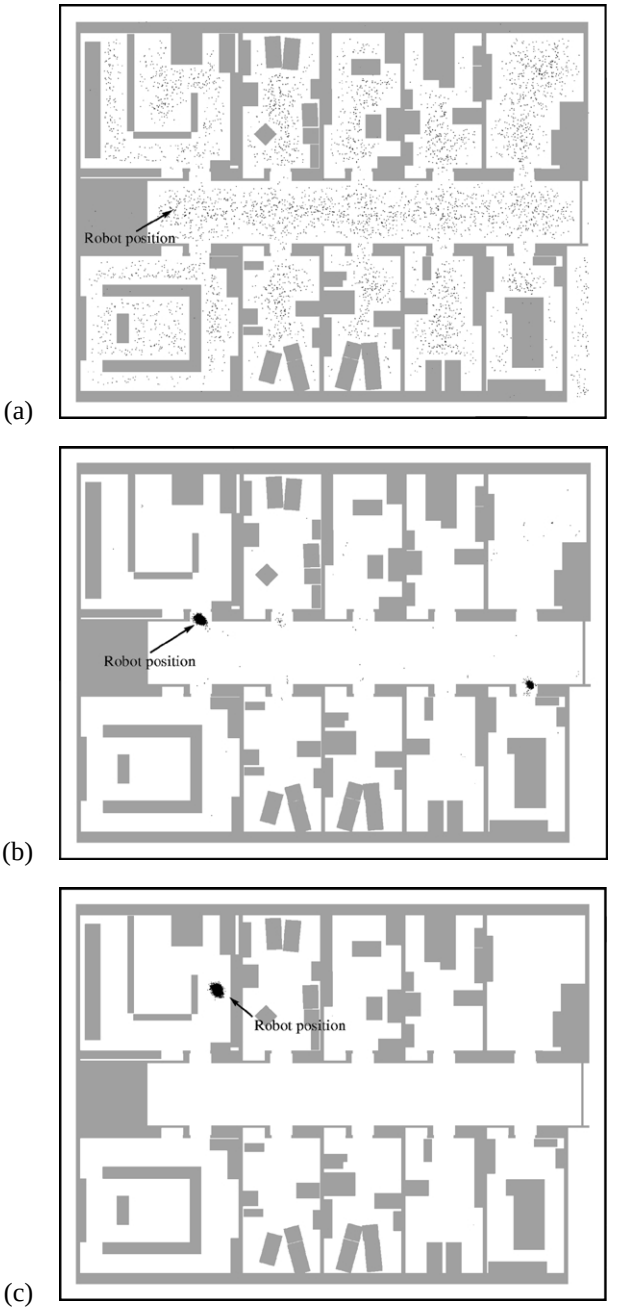
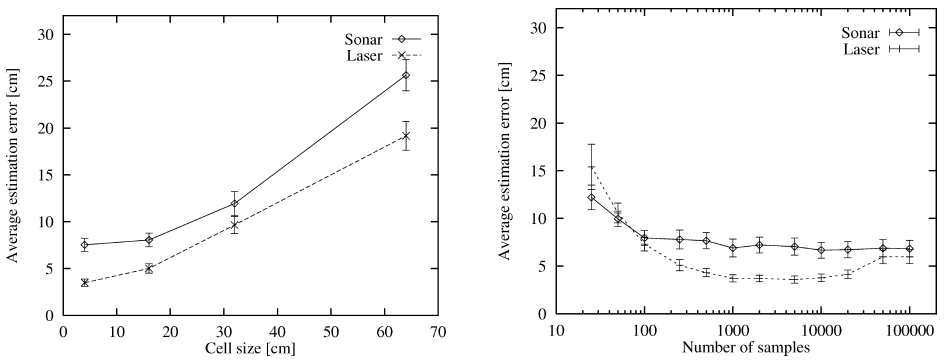


Fig. 6. Global localization of a mobile robot using MCL (10,000 samples): (a) initial particle set, uniformly distributed (projected into 2D). (b) Particles after approximately 2 meters of robot motion. Due to environment symmetry, most particles are centered around two locations. (c) Particle set after moving into a room, thereby breaking the symmetry. These experiments were carried out with an RWI B21 robot.



(a) (b)

Fig. 7. (a) Accuracy of grid-based Markov localization using different spatial resolutions. (b) Accuracy of MCL for different numbers of samples (log scale).

uncertain; hence the samples are spread uniformly through the free-space (projected into 2D). Fig. 6(b) shows the sample set after approximately 2 meters of robot motion, at which point MCL has disambiguated the robot's position up to a single symmetry. Finally, after another 2 meters of robot motion, the ambiguity is resolved, and the robot knows where it is. The majority of samples are now centered tightly around the correct position, as shown in Fig. 6(c).

Of particular interest shall be a comparison of MCL to an alternative localization algorithm capable of global mobile robot localization. In particular, we compared MCL to grid-based Markov localization, our previous best stochastic localization algorithm and one of the very few algorithms capable of localizing a robot globally [7,23]. The grid-based localization algorithm relies on a fine-grained piecewise constant approximation for the belief *Bel*, using otherwise identical sensor and motion models. The fact that our implementation employs identical sensor and motion models and is capable of processing the same data greatly facilitates the comparison. Fig. 7(a) plots the localization accuracy for grid-based localization as a function of the grid resolution. Notice that the results in Fig. 7(a) were not generated in real-time. As shown there, the accuracy increases with the resolution of the grid, both for sonar (solid line) and for laser data (dashed line). However, grid sizes below 8 cm do not permit updating in real-time in the specific testing environment, even when highly efficient selective update schemes are applied [24]. Results for MCL with fixed sample set sizes are shown in Fig. 7(b). These results have been generated using real-time conditions, where large sample sizes ($> 1{,}000$ samples) result in loss of sensor data due to time constraints. Here very small sample sets are disadvantageous, since they infer too large an error in the approximation. Large sample sets are also disadvantageous, since processing them requires too much time and fewer sensor items can be processed in real-time. The "optimal" sample set size, according to Fig. 7(b), is somewhere between 1,000 and 5,000 samples. Grid-based localization, to reach the same level of accuracy, has to use grids with 4 cm resolution—which is infeasible given even our best computers.

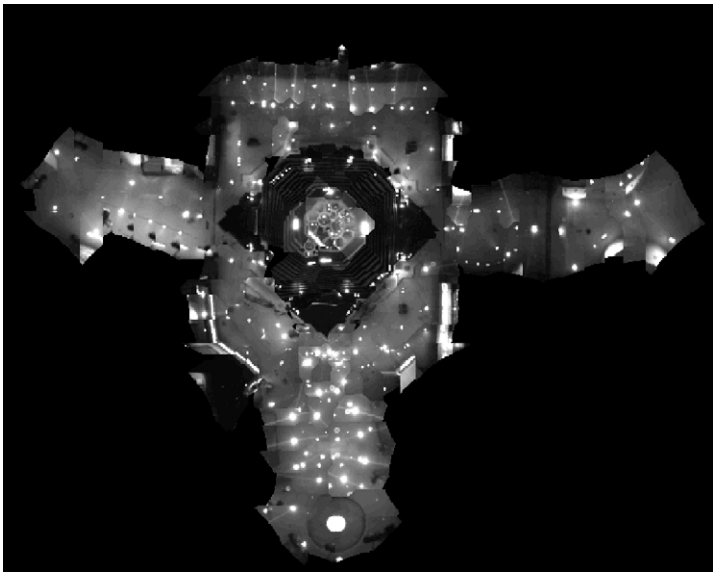


Fig. 8. Ceiling map of the National Museum of American History, which was used as the perceptual model in navigating with a vision sensor. The map was acquired using an RWI B18 robot.

2.7. Robot with upward-pointed camera

Similar results were obtained using a camera as the primary sensor for localization [12]. To test MCL under challenging real-world conditions, we evaluated it using data collected in a populated museum. During a two-week exhibition, our robot Minerva (Fig. 8) was employed as a tour-guide in the Smithsonian's Museum of Natural History, during which it traversed more than 44 km [71]. To aid localization, Minerva is equipped with a camera pointed towards the ceiling. Fig. 8 shows a mosaic of the museum's ceiling. Since the ceiling height is unknown, only the center region in the camera image is used for localization.

This data set is among the most difficult in our possession, as the robot traveled with speeds of up to 163 cm/sec. Whenever it entered or left the carpeted area in the center of the museum, it crossed a 2 cm bump which introduced significant errors in the robot's odometry. When only using vision information, grid-based localization fatally failed to track the robot. This is because the enormous computational overhead makes it impossible to incorporate sufficiently many images. MCL, however, succeeded in globally localizing the robot and tracking the robot's position in this specific data set. Fig. 9 shows a sequence of belief distributions during localization using MCL. However, as we will see below, in a second data sequence recorded during the museum's opening hours MCL would have failed to localize the robot if only vision information had been used—of course, in the actual exhibit [5] the laser was instrumental for localization.

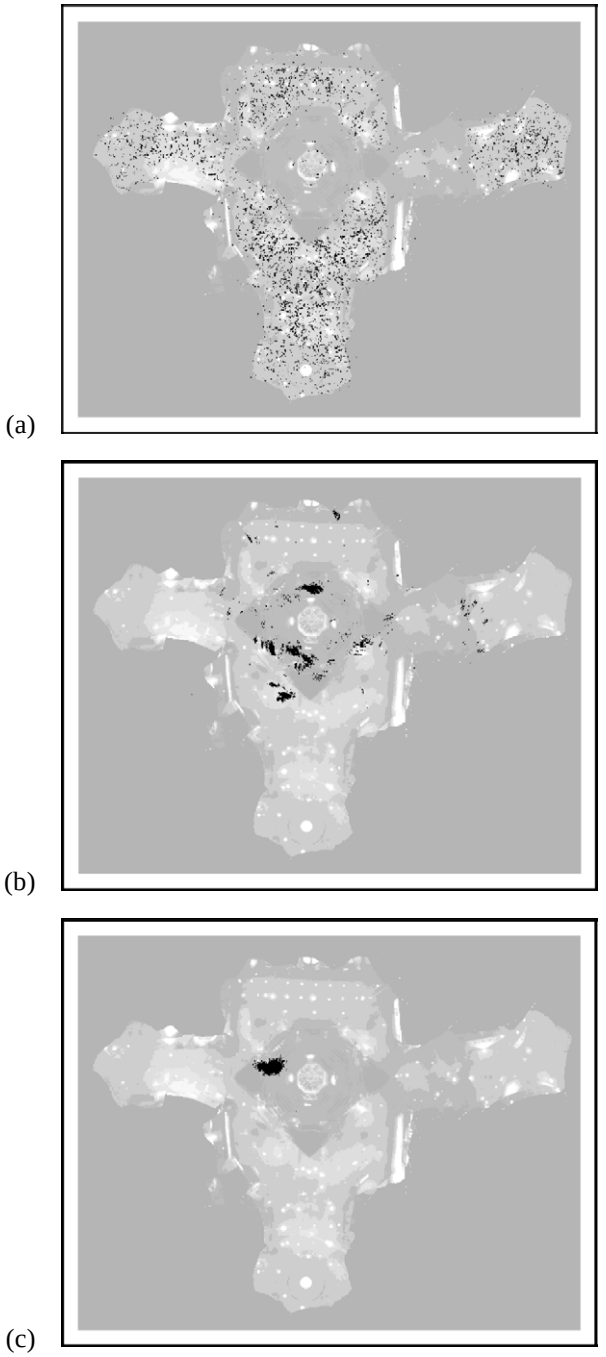


Fig. 9. Global localization of a mobile robot using a camera pointed at the ceiling. This experiment has been carried out with the Minerva tour-guide robot, an RWI B18 robot.

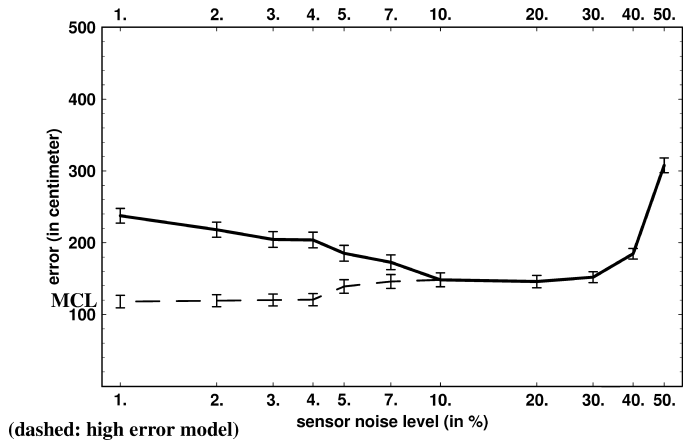


Fig. 10. Solid curve: error of MCL after 100 steps, as a function of the sensor noise. 95% confidence intervals are indicated by the bars. Notice that this function is *not* monotonic, as one might expect. Dashed curve: Same experiment with high-error model. These results were obtained through simulation.

3. Limitations of MCL

As noticed by several authors [17,46,58], the basic particle filter performs poorly if the proposal distribution, which is used to generate samples, places not enough samples in regions where the desired posterior $Bel(x_t)$ is large.

This problem has indeed practical importance in the context of MCL, as the following example illustrates. The solid curve in Fig. 10 shows, for our object localization example, the accuracy MCL achieves after 100 steps, using $m = 1{,}000$ samples. These results were obtained in simulation, enabling us to vary the amount of perceptual noise from 50% (on the right) to 1% (on the left). It appears that MCL works best for 10% to 20% perceptual noise. The degradation of performance towards the right, when there is a lot of noise, hardly surprises. The less accurate a sensor, the larger an error one should expect. However, MCL also performs poorly when the noise level is too small. In other words, MCL with accurate sensors may perform *worse* than MCL with inaccurate sensors. At first glance, this finding may appear to be a bit counter-intuitive in that it suggests that MCL only works well in specific situations, namely those where the sensors possess the "right" amount of noise.

Fig. 11 depicts an example run for highly accurate sensors in which MCL fails. When the object is first sighted, none of the samples is close enough to the object's true position. As a consequence, MCL gradually removes all samples with the exception of those located in the simulated robot's "dead spot", which is the center of its circular trajectory. Clearly, localization fails in this example, with a final error of 394 cm. Unfortunately, the more accurate the sensors, the smaller the support of $p(o \mid x)$, hence the more likely this problem occurs.

There are, of course, fixes. At first glance, one might add artificial noise to the sensor readings. A more sensible strategy would be to use a perceptual model $p(o_t \mid x_t)$ that overestimates the actual sensor noise, instead of adding noise to the sensor measurements. In fact, such a strategy, which has been adopted in [21], partially alleviates the problem:

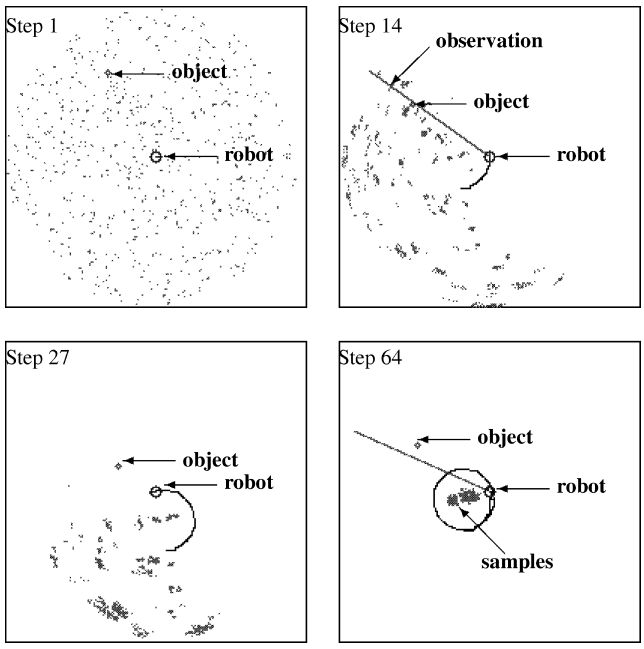


Fig. 11. Unsuccessful localization sequence for our robot simulation of object localization. The final error in this case is 394 cm.

The dashed curve in Fig. 10 shows the accuracy if the error model assumes a fixed 10% noise (shown there only for smaller "true" error rates). While the performance is better, this is hardly a fix. The overly pessimistic sensor model is inaccurate, throwing away precious information in the sensor readings. In fact, the resulting belief is not any longer a posterior, even if infinitely many samples were used. Other fixes include the addition of *random* samples into the posterior [21], and the generation of samples at locations that are consistent with the sensor readings [43]—a strategy that is similar to Mixture-MCL below but without proper sample weighting is mathematically questionable. While these approaches have shown superior performance over strict MCL in certain settings, neither of them can be expected to converge to the true posterior as the sample set size goes to infinity.

To analyze the problem more thoroughly, we first notice that the true goal of Bayes filtering is to calculate the product distribution specified in Eq. (13). Thus, the optimal proposal distribution would be this product distribution. However, sampling from this distribution directly is too difficult. As noticed above, MCL samples instead from the proposal distribution q_t defined in Eq. (12), and uses the importance factors (15) to account for the difference. It is well known from the statistical literature [17,46,58,69] that the divergence between (12) and (13) determines the convergence speed. This difference is accounted by the perceptual density $p(o_t \mid x_t)$: If the sensors are entirely uninformative, this distribution is flat and (12) is equivalent to (13). For low-noise sensors, however, $p(o_t \mid x_t)$ is typically quite narrow, hence MCL converges slowly. Thus, the error in

Fig. 10 is in fact caused by two different types of errors: one arising from the limitation of the sensor data (= noise), and one that arises from the mismatch of (12) and (13) in MCL. As we will show in this article, an alternative version of MCL exists that practically eliminates the second error source, thereby enhancing the accuracy and robustness of the approach.

4. Mixture-MCL

4.1. The dual of MCL

We will now derive an alternative version of MCL, called *dual Monte Carlo localization*. This algorithm will ultimately lead to the main algorithm proposed in this article, the *Mixture-MCL* algorithm.

The key idea of the dual is to "invert" the sampling process, by exchanging the roles of the proposal distribution and the importance factors in MCL. More specifically, dual MCL generates samples of the state $x_t^{(i)}$ by virtue of the following proposal distribution:

$$\bar{q}_t = \frac{p(o_t \mid x_t)}{\pi(o_t)} \quad \text{with } \pi(o_t) = \int p(o_t \mid x_t)\, \mathrm{d}x_t. \tag{16}$$

Here the normalizer, $\pi(o_t)$, is assumed to be finite, which indeed is the case for mobile robot localization in environments of bounded size. Dual MCL can be viewed as the logical inverse of the sampling in regular MCL: Rather than guessing the state $x_t^{(i)}$ and then using the most recent observation to adjust the importance of a guess, dual MCL guesses states corresponding to the most recent observation, and adjusts the importance factor in accordance with the prior belief $Bel(x_{t-1})$. Consequently, the dual proposal distribution possesses complimentary strengths and weaknesses: while it is ideal for highly accurate sensors, its performance is negatively affected by measurement noise. The key advantage of dual MCL is that when the distribution of $p(o \mid x)$ is narrow—which is the case for low-noise sensors—dual sampling can be more effective than conventional MCL.

4.2. Importance factors

We will now provide three alternative ways to calculate the importance factors for $\bar{q}_t$. The first is mathematically the most elegant, but for reasons detailed below it is not well suited for mobile robot localization. The other two require an additional smoothing step. Both work well for mobile robot localization.

4.2.1. Approach 1

(Proposed by Arnaud Doucet, personal communication.) The idea is to draw random pairs $\langle x_t^{(i)}, x_{t-1}^{(i)} \rangle$ by sampling $x_t^{(i)}$ as described above, and $x_{t-1}^{(i)}$ by drawing from $Bel(x_{t-1})$. Obviously, the combined proposal distribution is then given by

$$\bar{q}_{1,t} = \frac{p(o_t \mid x_t)}{\pi(o_t)} \times Bel(x_{t-1}) \tag{17}$$

and the importance factors are given by

$$w^{(i)} = p\left(x_t^{(i)} \mid x_{t-1}^{(i)}, a_{t-1}\right). \tag{18}$$

To see the latter, we notice that the importance factor is the quotient of the target distribution (13) and the proposal distribution (17), up to a scaling factor:

$$\begin{aligned}&\left[\frac{p(o_t \mid x_t^{(i)})}{\pi(o_t)}\, Bel(x_{t-1}^{(i)})\right]^{-1} \eta p\left(o_t \mid x_t^{(i)}\right) p\left(x_t^{(i)} \mid x_{t-1}^{(i)}, a_{t-1}\right) Bel\left(x_{t-1}^{(i)}\right)\\ &\quad = \eta p\left(x_t^{(i)} \mid x_{t-1}^{(i)}, a_{t-1}\right)\pi(o_t) \propto p\left(x_t^{(i)} \mid x_{t-1}^{(i)}, a_{t-1}\right).\end{aligned} \tag{19}$$

This approach is mathematically more elegant than the two alternatives described below, in that it avoids the need to transform sample sets into densities (which will be the case below). However, in the context of global mobile robot localization, the importance factor $p(x_t^{(i)} \mid a_{t-1}, x_{t-1}^{(i)})$ will be zero (or vanishingly small) for many pose pairs $\langle x_t^{(i)}, x_{t-1}^{(i)}\rangle$. This is because it is unlikely that *independently* drawn poses $x_t^{(i)}$ and $x_{t-1}^{(i)}$ are "consistent" with the action a_{t-1} under the motion model. We have therefore not implemented this approach. However, for other estimation problems using particle filters it might work well, which is the reason why it is described in this article.

4.2.2. Approach 2

The second approach seeks to calculate importance factors for samples $x_t^{(i)}$ more directly. This approach requires a separate forward phase, in which an auxiliary density function is constructed that is subsequently used to calculate importance factors.

In the forward phase, our approach generates samples $x_{t-1}^{(j)} \sim Bel(x_{t-1})$ and then $x_t^{(j)} \sim p(x_t \mid x_{t-1}^{(j)}, a_{t-1})$. As is easily seen, each resulting sample $x_t^{(j)}$ is distributed according to $p(x_t \mid a_{t-1}, d_{0\ldots t-1})$. Together, they represent the robot's belief at time t *before* incorporating the sensor measurement o_t. Belief distributions of this type are often referred to as *predictive distributions* in the Kalman filtering literature [51], since they represent the prediction of state x_t based on the action a_{t-1} and previous data, but before incorporating the sensor measurement at time t. To use this distribution for calculating importance factors, the 'trick' is to transform the samples $x_t^{(j)}$ into a kernel density tree (kd-tree) [3,53] that represents the predictive distribution $p(x_t \mid a_{t-1}, d_{0\ldots t-1})$ in closed form. Using this tree, we can now calculate $p(x_t^{(i)} \mid a_{t-1}, d_{0\ldots t-1})$ for any sample $x_t^{(i)}$ generated by the dual sampler.

We now have all pieces together to define the second version of the dual sampler. Similar to the previous approach, samples $x_t^{(i)}$ are generated according to the proposal distribution $\bar{q}_{2,t} = \bar{q}_t$ as defined in Eq. (16). However, instead of sampling from $Bel(x_{t-1})$, this approach directly assigns importance factors to samples $x_t^{(i)}$ using the kd-tree that represents

$$w^{(i)} = p\left(x_t^{(i)} \mid a_{t-1}, d_{0\ldots t-1}\right). \tag{20}$$

To verify the correctness of this approach, we notice that the quotient of the Bayes filter target distribution (9) and the proposal distribution (16) is indeed proportional to the importance factors $w^{(i)}$.

$$\left[\frac{p(o_t \mid x_t^{(i)})}{\pi(o_t)}\right]^{-1} \eta p\bigl(o_t \mid x_t^{(i)}\bigr) \int p\bigl(x_t^{(i)} \mid x_{t-1}, a_{t-1}\bigr) Bel(x_{t-1})\,\mathrm{d}x_{t-1}$$
$$= \left[\frac{p(o_t \mid x_t^{(i)})}{\pi(o_t)}\right]^{-1} \eta p\bigl(o_t \mid x_t^{(i)}\bigr) p\bigl(x_t^{(i)} \mid a_{t-1}, d_{0\ldots t-1}\bigr)$$
$$= \eta\pi(o_t)\, p\bigl(x_t^{(i)} \mid a_{t-1}, d_{0\ldots t-1}\bigr). \tag{21}$$

In comparison to the first approach discussed here, this approach avoids the danger of generating pairs of poses $\langle x_t^{(i)}, x_{t-1}^{(i)}\rangle$ for which $w^{(i)} = 0$. On the other hand, it involves an explicit forward sampling phase that requires the induction of a probability density function from samples using a kd-tree. The induction step smoothes the resulting density, which reduces the variance of the estimation. However, the primary role of converting samples into kd-trees is that it facilitates the calculation of the importance weights.

4.2.3. Approach 3

The third approach avoids the explicit forward-sampling phase of the second approach, but it tends to generate smaller importance factors. In particular, the third approach transforms the initial belief $Bel(x_{t-1})$ into a kd-tree, very much like in the forward phase of the second approach. For each sample $x_t^{(i)} \sim \bar{q}_t$, one now draws a sample $x_{t-1}^{(i)}$ from the distribution

$$\frac{p(x_t^{(i)} \mid x_{t-1}, a_{t-1})}{\pi(x_t^{(i)} \mid a_{t-1})}, \tag{22}$$

where

$$\pi\bigl(x_t^{(i)} \mid a_{t-1}\bigr) = \int p\bigl(x_t^{(i)} \mid x_{t-1}, a_{t-1}\bigr)\,\mathrm{d}x_{t-1}. \tag{23}$$

As above, we assume that the integral $\pi(x_t^{(i)} \mid a_{t-1})$ is finite, which is trivially the case in the context of mobile robot localization.

In other words, our approach projects $x_t^{(i)}$ back to a possible predecessor pose $x_{t-1}^{(i)}$. Consequently, the pair of poses $\langle x_t^{(i)}, x_{t-1}^{(i)}\rangle$ is distributed according to the proposal distribution

$$\bar{q}_{3,t} = \frac{p(o_t \mid x_t)}{\pi(o_t)} \times \frac{p(x_t \mid x_{t-1}, a_{t-1})}{\pi(x_t \mid a_{t-1})}, \tag{24}$$

which gives rise to the following importance factor

$$w^{(i)} = Bel\bigl(x_{t-1}^{(i)}\bigr). \tag{25}$$

To see, we notice that $w^{(i)}$ is proportional to the following quotient of the target distribution (13) and the proposal distribution $\bar{q}_{3,t}$ specified in (24):

$$\left[\frac{p(o_t \mid x_t^{(i)})}{\pi(o_t)} \frac{p(x_t^{(i)} \mid x_{t-1}^{(i)}, a_{t-1})}{\pi(x_t^{(i)} \mid a_{t-1})}\right]^{-1} \eta p\bigl(o_t \mid x_t^{(i)}\bigr) p\bigl(x_t^{(i)} \mid x_{t-1}^{(i)}, a_{t-1}\bigr) Bel\bigl(x_{t-1}^{(i)}\bigr)$$
$$= \eta\pi(o_t)\pi\bigl(x_t^{(i)} \mid a_{t-1}\bigr) Bel\bigl(x_{t-1}^{(i)}\bigr). \tag{26}$$

Table 2
The dual MCL algorithm (third approach)

Algorithm dual_MCL_3(X, a, o)
$X' = \emptyset$
generate kd-tree b from X
for $i = 0$ to m do
generate random $x' \sim p(x' \mid o)/\pi(o)$
generate random $x \sim p(x' \mid \bar{a}, x)/\pi(x' \mid a)$
$w' = b(x)$
add $\langle x', w' \rangle$ to X'
endfor
normalize the importance factors w' in X'
return X'

When calculating $w^{(i)}$, the term $Bel(x_{t-1}^{(i)})$ is calculated using the kd-tree representing this belief density. In our derivation, we silently assumed that the term $\pi(x_t^{(i)} \mid a_{t-1})$ is a constant, hence can be omitted. This is indeed a very good approximation for mobile robot localization, although in the general case of dynamical systems this assumption may not be valid. Table 2 shows the algorithm that implements this specific version of dual MCL.

4.2.4. Performance

The reader should notice that all three approaches require a method for sampling poses from observations according to $\bar{q}_t$—which can be non-trivial in mobile robot applications. The first approach is the easiest to implement and mathematically most straightforward. However, as noted above, we suspect that it will be inefficient for mobile robot localization. The two other approaches rely on a density estimation method (such as kd-trees). The third also requires a method for sampling poses backwards in time, which further complicates its implementation. However, the superior results given below may well make this additional work (i.e., implementing the dual) worthwhile.

Unfortunately, dual MCL alone is insufficient for localization. Fig. 12 depicts the performance of dual MCL using the third approach, under the same conditions that led to the MCL results shown in Fig. 10. In both figures, the horizontal axis depicts the amount of noise in perception, and the vertical axis depicts the error in centimeters, averaged over 1,000 independent runs. Two things are remarkable: First, the accuracy is now *monotonic* in perceptual noise: More accurate sensors give better results. Second, however, the overall performance is much poorer than that of conventional MCL. The poor performance of dual MCL is due to the fact that *erroneous* sensor measurements have a devastating effect on the estimated belief, since almost all samples are generated at the "wrong" place.

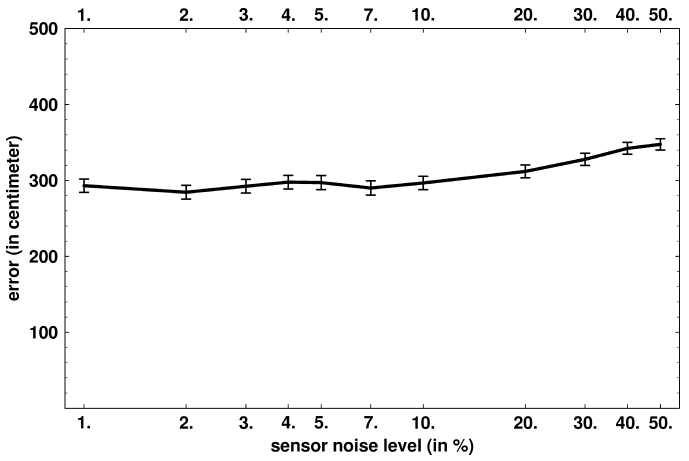


Fig. 12. Error of dual MCL as a function of the sensor noise. The error appears to increase monotonically with the sensor noise, but the overall error level is high. Compare this graph to Fig. 10. These results were obtained through simulation.

4.3. The Mixture-MCL algorithm

In many practical situations, plain MCL will work sufficiently well. However, as the experimental results suggest, neither version of MCL alone—the plain MCL algorithm and its dual—delivers satisfactory performance in certain situations. The plain MCL algorithm fails if the perceptual likelihood is too peaked. The dual MCL only considers the most recent sensor measurement, hence is prone to failure when the sensors fail, regardless of which of the three alternatives is used to calculate importance factors. However, each approach has complimentary strengths and weaknesses, suggesting that a combination of both might yield superior performance.

Mixture-MCL, the final algorithm described in this paper, merges both approaches. In Mixture-MCL, samples are generated by both plain MCL, and (one of) its duals, with a mixing rate of ϕ (with $0 \leqslant \phi \leqslant 1$). This is achieved by generating each sample with probability $1 - \phi$ using standard MCL, and with probability ϕ using a dual.

Table 3 states the Mixture-MCL algorithm, using the third variant for calculating importance factors in the dual. As is easy to be seen, the Mixture-MCL algorithm combines the MCL algorithms in Table 1 with the dual algorithm in Table 2, using the (probabilistic) mixing ratio ϕ. With probability $1 - \phi$, MCL's sampling mechanisms is used to generate a new sample. Otherwise, samples are generated via the dual. Notice that both types of samples are added to the same sample set *before* normalization. To make the two types of importance factors compatible, it is necessary to multiply the importance factors of the dual samples by $\pi(o)\pi(x' \mid a)$, when compared to the algorithm stated in Table 2. To see why, we notice that Eq. (15) suggests that the importance factors for plain MCL should be $\eta p(o \mid x)$. Similarly, the importance factors of the third version of the dual were derived as $\eta\pi(o)\pi(x' \mid a)b(x)$, according to Eq. (26), where b denotes the tree generated from the corresponding belief *Bel*. Since the constant η occurs in both of versions of MCL, it

Table 3
The Mixture-MCL algorithm, here using the third variant of dual MCL (see Table 2)

Algorithm MixtureMCL(X, a, o)

$X' = \emptyset$
generate kd-tree b from X
for $i = 0$ to m do
  with probability $1 - \phi$ do
    generate random x from X according to $w_1, \ldots, w_m$
    generate random $x' \sim p(x' \mid a, x)$
    $w' = p(o \mid x')$
    add $\langle x', w' \rangle$ to X'
  else do
    generate random $x' \sim p(x' \mid o)/\pi(o)$
    generate random $x \sim p(x' \mid \bar{a}, x)/\pi(x' \mid a)$
    $w' = \pi(o)\pi(x' \mid a)b(x)$
    add $\langle x', w' \rangle$ to X'
  endif
endfor
normalize the importance factors w' in X'
return X'

can safely be omitted. Thus, an appropriate importance factor for samples generated by regular MCL is $p(o \mid x')$, and $\pi(o)\pi(x' \mid a)b(x)$ for samples generated by its dual. In our implementation, we approximate $\pi(o)\pi(x' \mid a)$ by a constant.

Fig. 13 shows results obtained in simulation. Shown there are performance results of Mixture-MCL, using the second (thick line) and third (thin line) variant of the dual for calculating importance factors. All experiments use a fixed mixing ratio $\phi = 0.1$, and are averaged over 1,000 independent experiments per data point. A comparison with Fig. 10 suggests that Mixture-MCL is vastly superior to regular MCL, and in certain cases reduces the error by more than an order of magnitude. These results have been obtained with the third method for calculating importance factors. In our simulation experiments, we found that the second approach yields slightly worse results, but the difference was not significant at the 95% confidence level.

5. Sampling from the dual proposal distribution

For some sensors, sampling from the dual proposal distribution $\bar{q}$ can be far from trivial. For example, if the mobile robot uses proximity sensors and a metric map as described in Section 2.2, sampling from the inverse, $\bar{q}$ is *not* straightforward. The reader may recall that Section 2.2 outlines a closed-form routine for computing the forward model $p(o \mid x)$, which accepts both the pose x and an observation o as an input. However, dual MCL

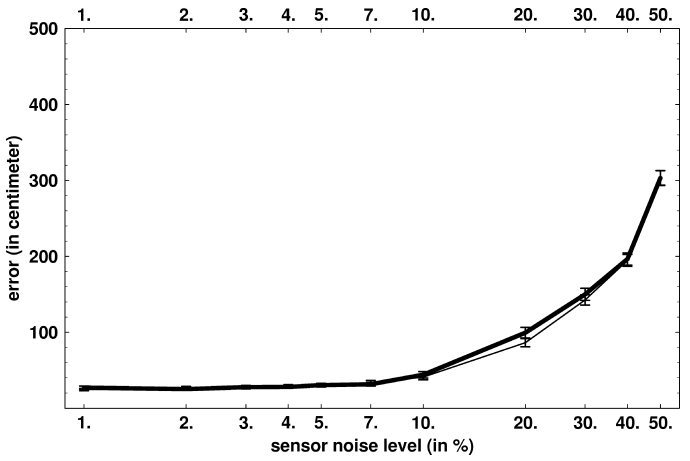


Fig. 13. Error of Mixture-MCL, combining regular and dual importance sampling. Mixture-MCL outperforms both components by a large margin, and its error is monotonic in the sensor noise. Thick line: second approach, thin line: third approach for dual sampling. Compare this graph to Figs. 10 and 12. These results were obtained through simulation.

requires us to sample poses x from a density proportional to $p(o \mid x)$, given just the observation o as an input. In other words, we are in need of a routine that accepts a range scan o as input, and which generates poses x.

The key idea here is to *"learn"* a sampling model of the joint distribution $p(o, x)$ from data, such that samples of the desired proposal distribution can be generated with ease. The specific representation chosen here is again a set of kd-trees, each of which models $p(o, x)$ for a subset of "similar" observations o, and each of which recursively partitions the space of all poses in a way that makes it easy to sample from $\bar{q} = p(o \mid x)/\pi(o)$.

The data used to construct the trees are samples $\langle x, o\rangle$ of poses x and observations o that are distributed according to the joint distribution, $p(x, o)$. There are two ways to sample from the joint: (1) synthetically, using the existing probabilistic models, and (2) using the physical robot to gather data (and the probabilistic model to augment such data).

- The *synthetic sampling* scheme is relatively straightforward. To generate a single sample, joint can be done in two cascaded sampling steps:
  (1) a pose x is sampled from a uniform distribution, and
  (2) for this pose, an observation is sampled according to $p(o \mid x)$.
  Sampling is repeated, until a sufficient number of samples has been generated (e.g., a million). Obviously, the resulting sample set represents the joint distribution $p(x, o)$.
- An alternative way to generate samples of the joint is to use *data collected by a physical robot*. This approach is preferable if one cannot easily sample from the perceptual model $p(o \mid x)$—as is actually the case in our existing software implementation (fixes such as rejection sampling [69] are inefficient). In general, sampling from the perceptual model is particularly difficult when using cameras, since this is a problem similar to the computer graphics problem. Luckily, sensor measurements o randomly collected by a robot are samples of $p(o)$, assuming that

the robot is placed randomly in its environment. However, robots are usually unable to measure their poses—otherwise there would not be a localization problem.

Luckily, importance sampling [62] offers a solution. The following routine generates samples from the desired joint distribution:

(1) Generate an observation o using a physical robot in a known environment with a random (but unknown) pose.
(2) Generate a large number of poses x according to a uniform distribution over the set of all robot poses.
(3) For each pose x, compute the *importance factor* $p(o \mid x)$ using the perceptual model described in Section 2.2.

Samples $\langle x, o \rangle$ generated using this approach, along with their importance factors $p(o \mid x)$, approximate the joint $p(o, x)$.

Equipped with samples representing the joint distribution, let us now turn our attention to learning trees that permit fast sampling from the dual proposal distribution $\bar{q}$. Again, there are multiple options to generate such trees. In our approach, the sensor measurements o are mapped into a low-dimensional feature vector. For laser range scans, a good set of features is the following three:

- The location of a sensor scan's center of gravity, relative to the robot's local coordinate system. The center of gravity is obtained by connecting the end-points of the individual range measurements and calculating the center of gravity of the enclosed area in 2D, relative to the robot's location. Its location is encoded in polar coordinates (two parameters).
- The *average* distance measurement, which is a single numerical parameter.

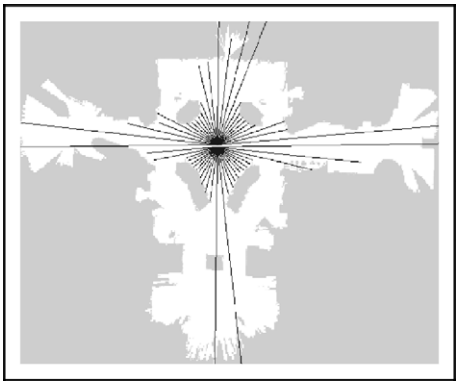
Together, these three values are treated like *sufficient statistics*, that is, we assume it suffices to know $f(o)$ to sample x, hence

$$\frac{p(o \mid x)}{\pi(o)} = \frac{p(f(o) \mid x)}{\pi(f(o))}. \tag{27}$$

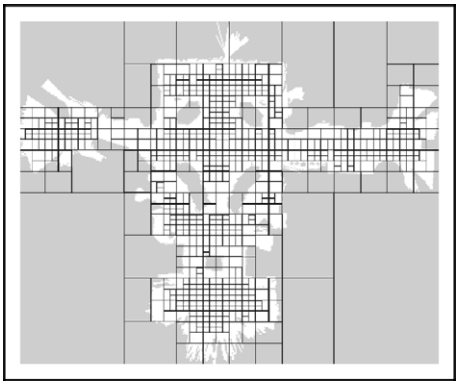
A discrete grid is then stipulated over these three values, and a kd-tree is grown for every (discrete) combination of the feature values $f(o)$. Each tree, thus, is conditioned on $f(o)$ (and hence on o). The depth of the tree depends on the total likelihood of a region in pose space: the more likely a pose given a specific observation $f(o)$, the deeper the tree (and hence the smaller the region covered by that leaf). Sampling from a tree is very efficient, since it only involves a small number of random coin flips.

Fig. 14 illustrates our sampling algorithm in practice. Shown in Fig. 14(a) is an example range scan along with an occupancy grid map [19,54] as described in Section 2.2. From this scan, our approach extracts the three features described above (center of gravity, average distance). Fig. 14(b) shows the tree that corresponds to these features, which partitions the state space recursively into small hyper-rectangular regions. Sampling from this tree yields sample sets like the one shown in Fig. 14(c).

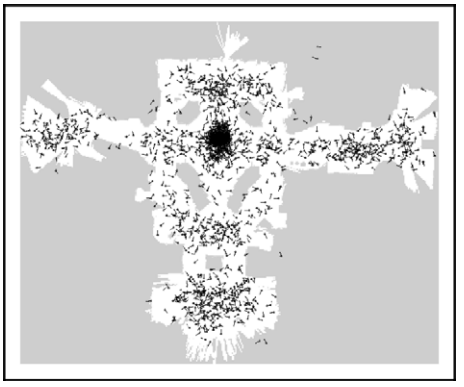
The tree constructed here can be built off-line, before robot operation. This is in contrast to the trees used to represent beliefs in Mixture-MCL, which must be built on-line after each belief update. This distinction is important when considering running times of the different variants of MCL. Because building a tree representing $p(o, x)$ is an off-line operation, it does not factor into the running time of the algorithm. It should be noted,



(a) laser scan and map



(b) tree for this scan



(c) samples of poses generated from tree

Fig. 14. Sampling from a distribution proportional to $p(x \mid o)$: (a) example range scan and map, (b) tree that partitions the state space for this scan, and (c) samples of poses x generated from the tree. The underlying data was collected by the Minerva data, an RWI B18 robot.

however, that the tree significantly increases the memory requirements of the approach. For example, the tree for generating samples from laser range scans in the museum environment requires approximately 80 MB of memory.

6. Experimental results

Systematic experimental results were conducted to evaluate the advantage of Mixture-MCL to regular MCL under a wide range of circumstances. The comparisons were carried out for a range of localization problems, with an emphasis on the more difficult global localization problem and the kidnapped robot problem. As above, real robot experiments were augments by systematic simulation results, where key parameters such as the amount of sensor noise could easily be controlled. When emulating global localization failures in the kidnapped robot problem, it is important that the tree used for calculating importance factors in the dual filter be non-zero everywhere. This was achieved by using a Dirichlet prior for estimating the probabilities in the leaves of the tree.

6.1. Simulation

Fig. 13 shows the performance of Mixture-MCL, under conditions that are otherwise identical to those in Fig. 10. As these results suggest, our new MCL algorithm outperforms both MCL and its dual by a large margin. At every single noise level, our new algorithm outperforms its alternatives by a factor that ranges from 1.07 (high noise level) to 9.7 (low noise level). For example, at a noise level of 1%, Mixture-MCL algorithm exhibits an average error of 24.6 cm, whereas MCL's error is 238 cm and that of dual MCL is 293 cm. In comparison, the average error with noise-free sensors and the optimal estimator is approximately 19.5 cm.

Mixture-MCL degrades gracefully to very small sample sets. Fig. 15 plots the error of conventional MCL (top curve) and Mixture-MCL (bottom curve) for different error levels, using $m = 50$ samples. With only 50 samples, regular MCL basically fails to track the robot's position in our simulation. Mixture-MCL exhibits excellent performance, and is only slightly inferior to $m = 1{,}000$ samples. Viewed differently, these findings suggest that Mixture-MCL is computationally an order of magnitude more efficient than conventional MCL.

Finally, Mixture-MCL tends to exhibit superior performance in the kidnapped robot problem. Fig. 16 shows the average localization error averaged over 1,000 simulation runs, as a function of time. The three different curves in that figure correspond to three different algorithms: MCL (thin curve), MCL with added random samples (dashed curve), and Mixture-MCL (thick curve). At time step 100, the simulated robot is kidnapped: it is tele-ported to a random pose without being told. As argued in the introduction, kidnapping is a way to test the ability of a localization algorithm to recover from catastrophic failures. As the results in Fig. 16 suggest, Mixture-MCL recovers faster than any alternative MCL algorithm, despite the fact that we optimized parameters such as the ratio of random samples beforehand. Regular MCL fails entirely to recover from the kidnapping, since it tends to lack samples at the new robot pose. The addition of random samples overcomes

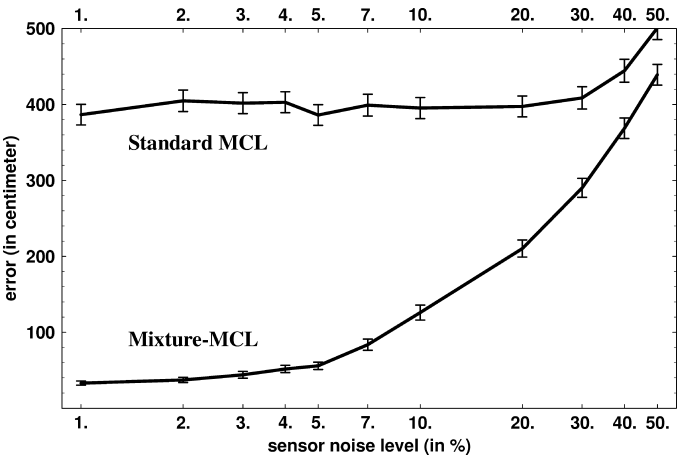


Fig. 15. Error of MCL (top curve) and Mixture-MCL (bottom curve) with 50 samples (instead of 1,000) for each belief state. These results were obtained through simulation.

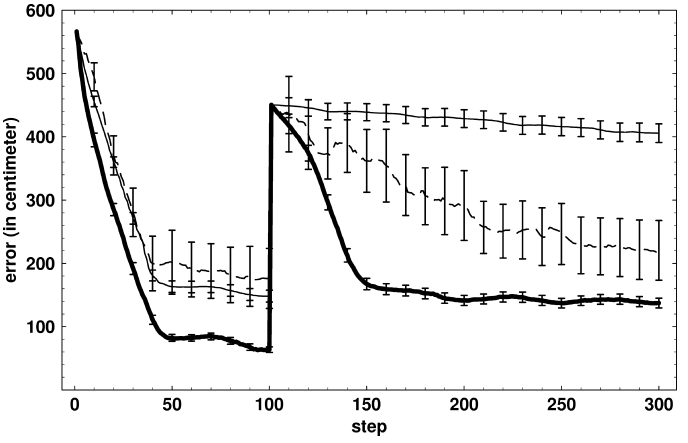


Fig. 16. Kidnapped robot problem: Localization error as a function of time, for three different approaches: MCL (thin curve), MCL with added random samples (dashed curve), and Mixture-MCL (thick curve). At time $t = 100$, the robot is tele-ported to a random pose without being told. As these results suggest, Mixture-MCL is most efficient in recovering from this incident. These results were obtained through simulation.

this problem, but is inefficient. Mixture-MCL places samples more thoughtfully, which increases its efficiency in recovering from kidnapping.

6.2. Robot with laser range finder

Mixture-MCL has also been evaluated using data recorded by Minerva. As outlined above, the data contains logs of odometry measurements and sensor scans taken by

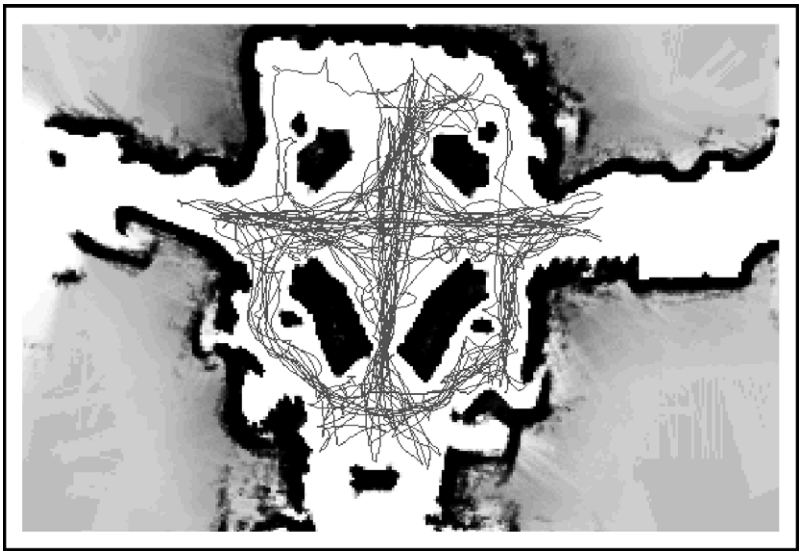


Fig. 17. Estimated path of the Minerva robot in the Smithsonian Museum of National History.

Minerva's two laser range-finders (see [22] for details). Fig. 17 shows part of the map of the museum and the path of the robot used for this evaluation.

As already reported in Section 2.4, conventional MCL reliably succeeds in localizing the robot. Thus, our attention here is to evaluate Mixture-MCL for the kidnapped robot problem. To do so, we repeatedly introduced errors into the odometry information. These errors made the robot lose track of its position with probability of 0.01 when advancing one meter.

Fig. 18 shows comparative results for our three different approaches. The error is measured by the percentage of time, during which the estimated position deviates by more than 2 meters from the reference position. Obviously, Mixture-MCL yields significantly better results, even if plain MCL is augmented by 5% random samples. Mixture-MCL reduces the error rate of localization by as much as 70% more than plain MCL; and 32% when compared to the case where plain MCL is augmented with uniform samples. These results are significant at the 95% confidence level.

6.3. Robot with upward-pointed camera

We also compared Mixture-MCL in the context of visual localization, using only camera imagery obtained with the robot Minerva during public museum hours [12]. Notice that this data set is *not* the same as the one used above; in particular, it contains an unexplained, unnaturally large odometry error, which occurred for unknown reasons. In this particular case, the odometry reported back by the robot's low-level motion controller jumped by a large mount. We suspect that the error was caused by an overflow in the low-level robot motion control software, which is inaccessible to us. Since we did not expect such an error, our software did not check for unnaturally large odometry readings and

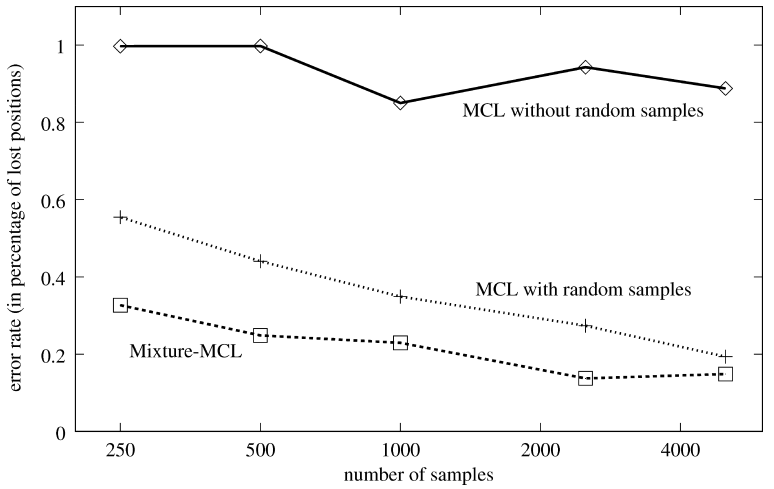
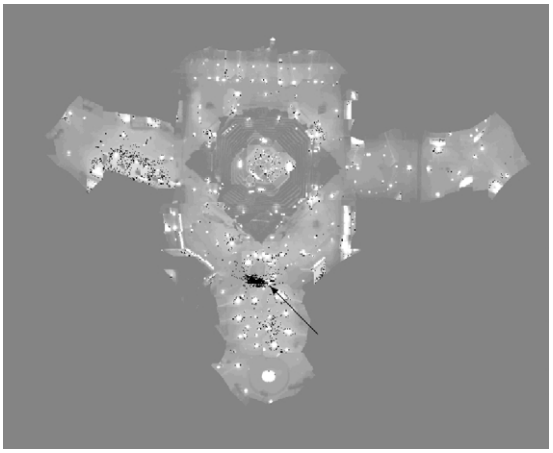


Fig. 18. Performance of conventional (top curve), conventional with random samples (middle curve) and mixture (bottom curve) MCL for the kidnapped robot problem in the Smithsonian museum. The error rate is measured in percentage of time during which the robot lost track of its position. These results were obtained using a physical robot.

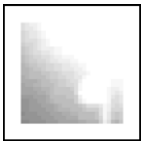
accepted it as if it was correct. Such an error, whatever its cause, induces a kidnapped robot problem. Moreover, the image sequence used for evaluation is of poor quality, as people often intentionally covered the camera with their hand and placed dirt on the lens.

Fig. 19 shows two sample sets (large images), superimposed on the ceiling mosaic, which have been generated by Mixture-MCL during localization. Arrows mark samples generated by the regular MCL sampler. Next to these diagrams, the center regions of the most recent camera images are shown (small diagrams), which are used for generating samples in the dual filter. In Fig. 19(a), the most recent image suggests that the robot is under a ceiling light. Consequently, the dual sampler generates samples close to light sources. In Fig. 19(b), the camera measurement is considerably dark, suggesting a location in the center octagon. Notice that we changed the brightness of the ceiling map to increase the visibility of the samples; the authentic ceiling map is shown in Fig. 8.

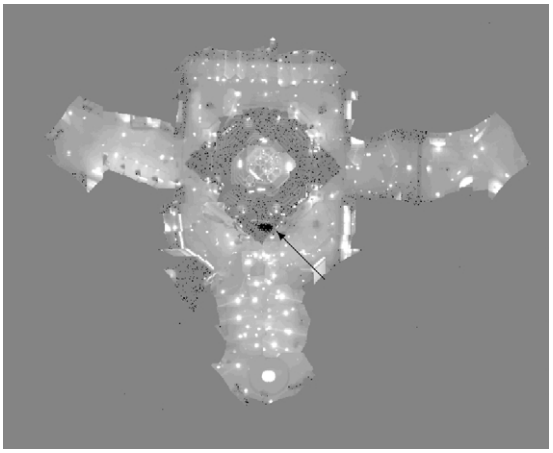
Fig. 20 depicts the localization error obtained when using vision only (calculated using the localization results from the laser as ground truth). The data covers a period of approximately 4,000 seconds, during which MCL processes a total of 20,740 images. After approximately 630 seconds, the aforementioned error in the robot's odometry leads to a loss of the position. As the two curves in Fig. 20 illustrate, regular MCL (dashed curve) is unable to recover from this event, whereas Mixture-MCL (solid curve) recovers. For this data set, MCL with added random sample performs similarly well as Mixture-MCL. These results are not statistically significant in that only a single run is considered, but they confirm our findings with laser range finders, indicating that Mixture-MCL is indeed a robust localization method.



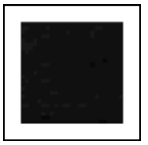
(a) sample set and ceiling mosaic



camera image (center region)



(b) sample set and ceiling mosaic



camera image (center region)

Fig. 19. Two sample sets generated by Mixture-MCL, along with the most recent camera image (center region of the image only). The arrows mark the samples generated using the conventional MCL sampler, whereas the other samples are generated by the dual. In (a), the most recent camera measurement suggests that our Minerva robot is near a ceiling light, whereas in (b) the measurement suggests a location in the center octagon.

6.4. Running times

A final issue addressed in our experiments concerns the running time of MCL algorithms. Clearly, the absolute time depends on a variety of factors, such as the number of samples, the nature of the probabilistic models and the sensor data, the amount of pre-processing required (e.g., for extracting features from camera images), and the underlying computer platform. The numbers discussed in this section shed some light onto the relative

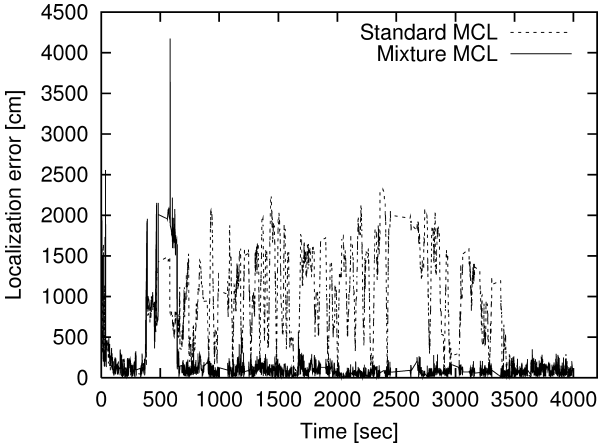
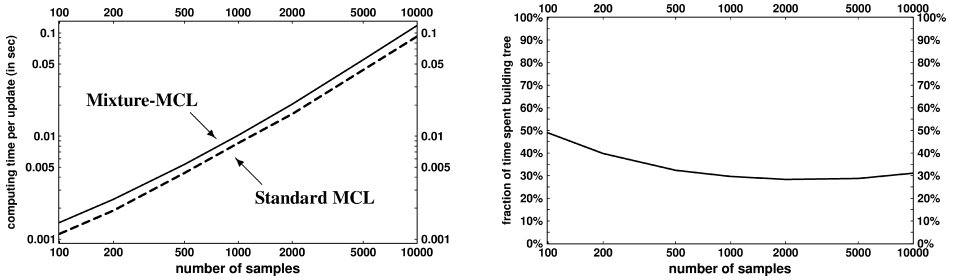


Fig. 20. Plain MCL (dashed curve) compared to Mixture-MCL (solid line). Shown here is the error for a 4,000-second episode of camera-based localization in the Smithsonian museum, using the robot Minerva.



(a) (b)

Fig. 21. (a) Time (in seconds on a 500 MHz Pentium PC) required for one full belief computation using regular MCL (dashed line) and Mixture-MCL (solid line), plotted as a function of the number of particles. (b) Percentage of time spent generating density trees in Mixture-MCL. All results are averaged over 1,000 updates using the robot simulator.

requirements of MCL and Mixture-MCL. They also illustrate that MCL methods can be implemented highly efficiently, requiring only a small fraction of the computational resources of a typical PC. All results reported here were obtained with a Pentium PC running at 500 MHz and equipped with sufficient RAM to hold all data in main memory.

Fig. 21(a) shows the time required for a full MCL update as a function of the number of samples, for our robot simulation. The solid line in Fig. 21(a) corresponds to regular MCL, whereas the dashed line shows the time required to run Mixture-MCL. Both coordinate axes are logarithmic. The time measurement is approximately linear, as one would expect when increasing the number of samples. With $m = 1{,}000$, regular MCL consumes on average 8.56×10^{-3} seconds per update. Mixture-MCL with mixing ratio $\phi = 0.1$ requires 1.02×10^{-2} seconds, which is 19.1% slower than regular MCL. This result appears to be invariant to the mixing ratio ϕ as long as $\phi > 0$. For $m = 1{,}000$, 29.7% of the total time

is spent on constructing the kd-tree. Fig. 21(b) plots the fraction of time spent building the tree in Mixture-MCL for different sample set sizes m. As can be seen there, the curve first decreases with increasing sample size (from 48.9% for $m = 100$ samples to 28.7% for $m = 5{,}000$). For $m = 10{,}000$ samples, the percentage increases slightly (to 31.1%), an increase that is statistically significant at the 95% level. The exact cause of this non-monotonic behavior is unknown to us; we attribute it to side effects of the PC architecture (e.g., computing with cache versus main memory).

The timing results of our physical robot implementation are similar. The running times for the MCL implementation using range data are shown in Fig. 22. This diagram shows the computation time on a 500 MHz Pentium PC, as a function of the sample set size, both plotted in logarithmic scale. The computation time is broken town into the two basic components: the integration of odometry measurements (curve marked with solid black squares), and the integration of range data (all other curves). For example, with $m = 1{,}000$ samples, both types of updates can be performed in less than five thousands of a second, which is approximately 40 times faster than sensor data arrives.

Fig. 22 shows four different timing graphs for integrating range measurements. These correspond to different sensors (laser versus sonar) and different implementations. The top two curves in Fig. 22 depict the computation time for a straightforward implementation of MCL. This implementation calculates the distance to the nearest obstacle *on-line*, while integrating sensor data. Two of the three bottom curves show the computation time for

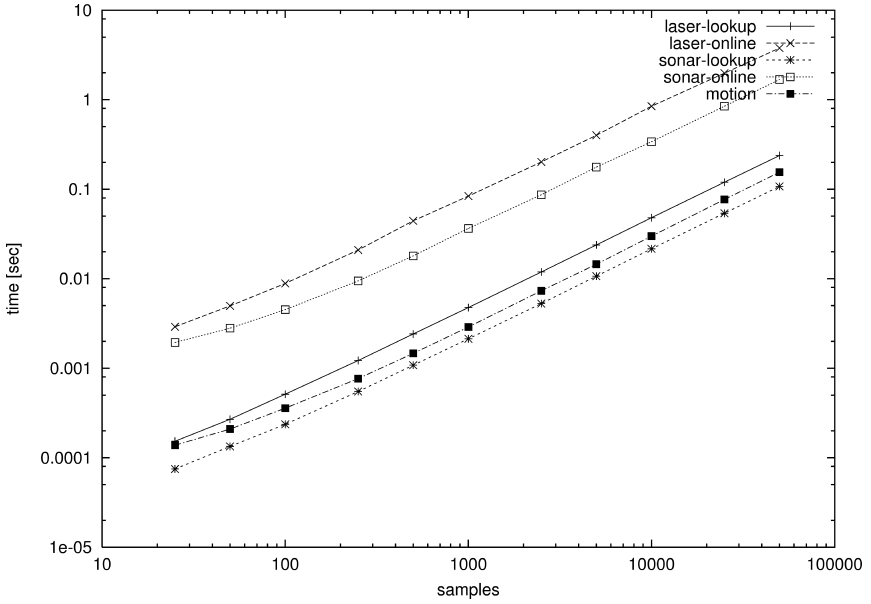


Fig. 22. Processor time required by our MCL implementation using range data. The results are broken down into the prediction step (bottom curve), where odometry data is processed by sampling new poses, and the observation step (top curve), where range data is incorporated into the importance factors. Notice that both axes are logarithmic. The bars indicate 95% confidence intervals.

an implementation where these distances are *pre-computed* and stored in a large table (see [24] for more details). By pre-calculating these distances, the integration of sensor measurements is sped up by a factor of 16.9. The downside of this technique is its memory requirement, which lies between 50 MB and 200 MB for indoor maps like the ones shown in this paper. The difference in computation time between laser and sonar data, which can be observed regardless of whether or not distances are pre-computed, stems from the fact that there are many more laser beams per scan than there are sonar beams. Our current software integrates 60 individual laser measurements for each laser scan, compared to 12 individual sonar measurements per sonar scan.

All these results were obtained for sample sets of fixed size. We notice that our physical robot implementation of MCL generates sample sets of variable sizes, driven by events. More specifically, our implementation generates samples until a new sensor measurement becomes available (up to a maximum number of samples). The advantage of such an implementation is its adaptivity to the available computational resources. Such resource-adaptive algorithms are sometimes referred to as *any-time algorithms* [11,76]. They have the advantage that when ported to a different computer (e.g., a new, faster PC), they can exploit the additional computational power without any modification to the program code. An argument in [5] emphasizes the use of resource-adaptive algorithms as a generic design principle of mobile robot software.

7. Related work

Mobile robot localization is a fundamental problem in mobile robotics, which has received considerable attention over the past decades [4,10,25,31,41,45,59,65,74]. As argued in the introduction of this article, the vast majority of work focuses on the position tracking problem, where errors are assumed to be small. Most approaches are incapable of recovering from localization failures, though methods exist for detecting such conditions. Usually, failures of the localization component require that a robot's position be entered manually.

Approaches that solve the global localization and the kidnapped robot problem are relatively recent, and they commonly rely on Bayes filtering with multi-modal density representations, just like MCL. A recent article [24] gives a comprehensive overview of algorithms for mobile robot localization with many references. As argued in the introduction of this article, there are several alternatives to the approach proposed here. Among the most prominent ones are probabilistic algorithms that use piecewise constant functions and Gaussian mixtures to represent the robot's belief. The former are known as Markov localization algorithms, and the latter are often implemented as multi-hypothesis Kalman filters. All of these approaches can be derived from the Bayes filter described in Section 2.1, which is also the mathematical basis of the various MCL algorithms presented in this article. Thus, all of these algorithms share the same mathematical basis.

Example for approximating the posterior using piecewise constant densities can be found in [7,24,30,36,40,50,55,56,66,70]. Many of these approaches approximate the belief using *topological* representations of robot environments. In such representations, the environment is decomposed into a small number of significant places, whose size

and location depends on the structure of the environment. The belief distribution is approximated by a finite distribution parameterized by these places, sometimes along with the heading direction. In [24], a variant is described that uses a fine-grained metric grid to represent the belief. Since the belief space is three-dimensional, the size of the grid is immense. Thus, [24] presents an approximate updating algorithm that restricts updates to a small subset of all grid cells that are deemed most relevant. This idea is carried further in [6], which proposes to use trees for representing beliefs. These trees represent probability densities with varying resolution, so that more likely regions are approximated more accurately (similar to the kd-trees generated from samples for dual MCL). All these approaches differ from the MCL family of algorithms in that they use parametric representations. They are difficult to implement if high accuracy is needed, but today's best implementations yield somewhat inferior performance, as suggested by the comparison in Section 2.6.

Localization algorithms based on the multi-hypothesis Kalman filter [1,2] represent beliefs using mixtures of Gaussians [9,34,60,61]. To calculate the covariance matrices of the individual Gaussian mixture components, the Kalman filtering approach linearizes the motion model and the perceptual model (see [35] for a recent non-linear extension of Kalman filters). It also assumes that errors in sensor measurements and robot motion are Gaussian. For most robot sensors, measurement noise is not Gaussian. Therefore, Kalman filtering algorithms usually do not use raw sensor data for localization. Instead, they extract features from which robot poses can be estimated with (assumed) Gaussian noise [9,34,60,61]. The literature suggests a range of methods to extract features, such as point features, line features, pairs of points, etc. Using features instead of the raw sensor data can be loss-free if the features are sufficient statistics of the sensor data relative to the problem of estimating poses. In practice, however, this is usually not the case, and significant information may be lost when going from raw data to features. Herein lies a primary difference to the MCL algorithms presented in this article, which can handle arbitrary noise models and are capable of using raw sensor data (e.g., laser range data) for localization.

Nevertheless, multi-hypothesis Kalman filters have been applied with great success to various versions of the localization problem, including position tracking and global localization [34,60,61]. Certain update steps in the multi-hypothesis Kalman filter can be leveraged across multiple Gaussians, which leads to an efficient implementation [61]. The basic update equations of these approaches can be shown to be hybrid versions of the Bayes filter with continuous and discrete components. Thus, these algorithms are imminently related to the MCL algorithms described here. From the conceptual point of view, Gaussian mixtures are similar to sample sets, with the key difference that Gaussians are continuous distributions, not just discrete samples, since they possess an associated covariance matrix. Two of the three versions of Mixture-MCL, for example, require a step where a kd-tree is generated from a sample set; such a set would not be necessary with the Gaussian representations, since mixtures of Gaussians are already continuous distributions. To keep the number of mixture components manageable in real-time, the approaches referenced above apply heuristics for terminating unlikely Gaussians and creating new ones when indicated by the sensor data. These heuristics are similar, but not identical, to the techniques proposed here. In MCL, unlikely samples are terminated probabilistically, as a side effect

of the sampling step. Mixture-MCL creates new hypotheses based on momentary sensor measurements, but it does so stochastically, and it considers the previous belief when determining the initial weight (probability) given to a new hypothesis. The significance of these differences are currently poorly understood. Generating a sample is generally faster than a Kalman filter update (which requires matrix inversion), but we suspect that more samples are needed to approximate a density than Gaussian mixtures.

Particle filters, as basic statistical tools, have become popular for tracking and position estimation in the last few years, as for example documented by a forthcoming book on this topic [18]. Recent research, has led to a range of variants of the basic particle filters. The poor performance of particle filtering in cases where the proposal distribution differs significantly from the target distribution has been observed by several authors, e.g., [17,39, 46,58]. Typical "fixes" involve the design of a different proposal distribution that places more weight on the tails of a distribution. In this light, Mixture-MCL can be viewed as one way to deal with this mismatch problem, one that works well for mobile robot localization.

Particle filters have been applied with great success to other estimation and tracking problems of practical importance. In computer vision, particle filters are commonly known as *condensation algorithm*, where they have been applied with remarkable success to visual tracking problems [32,33,48]. Their application to mobile robot localization has been proposed in [13,21] and since been adopted (and extended) by several other researchers [16,43]. In our own work, we recently extended the basic paradigm to collaborative localization for a whole team of mobile robots [22].

To the best of our knowledge, the idea of the dual particle filter proposed here and in [72] is new. Obviously, it works well in the context of mobile robot localization. While the aim of the article is to evaluate the Mixture-MCL algorithm in practice, it should be straightforward to devise a proof of convergence of all three versions of Mixture-MCL, assuming convergence of kd-trees. The idea of a dual is related to a recent article by Lenser and Veloso [43], who also propose to generate samples in accordance with the most recent sensor measurement. Like us, they evaluated their approach in the context of mobile robot localization. There are two main differences between their and our work: First, their approach generates samples that maximize the perceptual density $p(o \mid x)$, instead of sampling from $p(o \mid x)$. Second, and more importantly, their approach does not take past evidence into account when generating samples from sensor readings, that is, their approach does not adjust the importance factors of samples generated by the dual in accordance with $Bel(x_{t-1})$. Consequently, the resulting estimate does not approximate the posterior. For example, if the environment consists of disconnected components (e.g., rooms), such an approach can place non-zero likelihood behind walls that are physically impossible to traverse. Our approach relies on the same basic idea, but asymptotically approximates the desired posterior.

The idea of sampling from the sensor measurement (the "evidence") has also been proposed in the context of Bayes networks [29,57], in particular in the context of marginalization using Monte Carlo sampling. Under the name of "arc reversal", Kanazawa and colleagues [38] have proposed an efficient sampling algorithm that jump-starts samples at Bayes network nodes whose value is known, then propagating those samples throughout the network to obtain an estimate of the desired marginal distribution. This approach

is significantly more efficient than the importance sampler in Bayes networks (which follows the causality expressed by the Bayes network), for reasons that are identical to those given here. Our approach can be viewed as implementing this idea in the context of particle filtering, using somewhat different mathematical equations to account for the differences of Bayes networks and particle filtering. Also, our approach combines both sampling methodologies, which is essential for the superior performance of this approach.

8. Conclusion

This article introduced a new mobile robot localization algorithm, called Mixture Monte Carlo Localization. Mixture-MCL is a version of particle filters that combines a regular sampler with its dual. By combining both, our approach overcomes a range of limitations that currently exist for different versions of MCL, such as the inability to estimate posteriors for highly accurate sensors, poor degradation to small sample sets, and the ability to recover from unexpected large state changes (robot kidnapping).

Mixture-MCL possesses a range of unique advantages over previous localization algorithms capable of global localization from ambiguous features:

(1) *Efficiency.* Mixture-MCL inherits its computational efficiency from particle filters, which focus computational resources in areas that are most probable.
(2) *Versatility.* It also inherits from particle filters the ability to approximate a huge range of non-parametric densities, and to accommodate (almost) arbitrary non-linear system dynamics, sensor characteristics, and non-Gaussian noise. Often, the posterior is centered on a small subspace of the state space. Mixture-MCL does not require an explicit, parametric model of this subspace; instead, it models such subspaces implicitly by generating samples accordingly.
(3) *Resource adaptiveness.* Our implementation of Mixture-MCL is *any-time* [11, 76], in that the number of samples can be determined dynamically based on the available computational time between two consecutive sensor measurements. As a consequence, the software can be run on many different computer platforms, where it adapts to the available computational resources.
(4) *Robustness.* By mixing regular forward sampling with its dual, Mixture-MCL performs robustly under a range of circumstances, such as highly accurate sensors, robot kidnapping, and very small sample sets.

Extensive experimental results suggest that Mixture-MCL consistently outperforms MCL and related Markov localization algorithms.

While this article focused on the mobile robot localization problem, we conjecture that its basic algorithms transcend to a much broader range of state estimation problems for temporal dynamic systems. Bayes filters have been applied to estimation problems for decades, and a recent interest in Monte Carlo approximations [18,26] suggests that the probabilistic paradigm is well suited for a broad range of state estimation problems in noisy temporal domains. While this article has described the limitations of particle filtering in the context of mobile robot localization, we envision that many other estimation domains

might suffer similar problems that can be overcome by mixing particle filters with their duals.

Acknowledgements

The authors would like to thank Nando De Freitas and Arnaud Doucet, whose comments on a related paper substantially contributed to the presentation of the material. They also thank Scott Lenser and two anonymous reviewers for suggestions that helped improving this manuscript.

This research is sponsored by the National Science Foundation (and CAREER grant number IIS-9876136 and regular grant number IIS-9877033), and by DARPA-ATO via TACOM (contract number DAAE07-98-C-L032) and DARPA-ISO via Rome Labs (contract number F30602-98-2-0137), which is gratefully acknowledged. The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing official policies or endorsements, either expressed or implied, of the United States Government or any of the sponsoring institutions.

References

[1] Y. Bar-Shalom, T.E. Fortmann, Tracking and Data Association, Academic Press, New York, 1998.
[2] Y. Bar-Shalom, X.-R. Li, Estimation and Tracking: Principles, Techniques, and Software, YBS, Danvers, MA, 1998.
[3] J.L. Bentley, Multidimensional divide and conquer, Comm. ACM 23 (4) (1980) 214–229.
[4] J. Borenstein, B. Everett, L. Feng, Navigating Mobile Robots: Systems and Techniques, A.K. Peters, Wellesley, MA, 1996.
[5] W. Burgard, A.B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, S. Thrun, Experiences with an interactive museum tour-guide robot, Artificial Intelligence 114 (1–2) (1999) 3–55.
[6] W. Burgard, A. Derr, D. Fox, A.B. Cremers, Integrating global position estimation and position tracking for mobile robots: The dynamic markov localization approach, in: Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'98), Victoria, BC, 1998.
[7] W. Burgard, D. Fox, D. Hennig, T. Schmidt, Estimating the absolute position of a mobile robot using position probability grids, in: Proc. AAAI-96, Portland, OR, 1996.
[8] I.J. Cox, Blanche—An experiment in guidance and navigation of an autonomous robot vehicle, IEEE Transactions on Robotics and Automation 7 (2) (1991) 193–204.
[9] I.J. Cox, J.J. Leonard, Modeling a dynamic environment using a Bayesian multiple hypothesis approach, Artificial Intelligence 66 (1994) 311–344.
[10] I.J. Cox, G.T. Wilfong (Eds.), Autonomous Robot Vehicles, Springer, Berlin, 1990.
[11] T.L. Dean, M. Boddy, An analysis of time-dependent planning, in: Proc. AAAI-92, San Jose, CA, 1988, pp. 49–54.
[12] F. Dellaert, W. Burgard, D. Fox, S. Thrun, Using the condensation algorithm for robust, vision-based mobile robot localization, in: Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, 1999.
[13] F. Dellaert, D. Fox, W. Burgard, S. Thrun, Monte Carlo localization for mobile robots, in: Proc. IEEE International Conference on Robotics and Automation (ICRA-99), Detroit, MI, 1999.
[14] F. Dellaert, C. Thorpe, S. Thrun, Mosaicing a large number of widely dispersed, noisy, and distorted images: A Bayesian approach, Technical Report CMU-RI-TR-99-34, Carnegie Mellon University, Pittsburgh, PA, 1999.

[15] A.P. Dempster, A.N. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. Roy. Statist. Soc. Ser. B 39 (1) (1977) 1–38.
[16] J. Denzler, B. Heigl, H. Niemann, Combining computer graphics and computer vision for probabilistic self-localization, Internal Report, 1999.
[17] A. Doucet, On sequential simulation-based methods for bayesian filtering, Technical Report CUED/F-INFENG/TR 310, Cambridge University, Department of Engineering, Cambridge, UK, 1998.
[18] A. Doucet, J.F.G. de Freitas, N.J. Gordon (Eds.), Sequential Monte Carlo Methods In Practice, Springer, New York, 2001.
[19] A. Elfes, Occupancy grids: A probabilistic framework for robot perception and navigation, Ph.D. Thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 1989.
[20] S. Engelson, D. McDermott, Error correction in mobile robot map learning, in: Proc. 1992 IEEE International Conference on Robotics and Automation, Nice, France, 1992, pp. 2555–2560.
[21] D. Fox, W. Burgard, F. Dellaert, S. Thrun, Monte carlo localization: Efficient position estimation for mobile robots, in: Proc. AAAI-99, Orlando, FL, 1999.
[22] D. Fox, W. Burgard, H. Kruppa, S. Thrun, Collaborative multi-robot localization, Autonomous Robots 8 (3) (2000).
[23] D. Fox, W. Burgard, S. Thrun, Active Markov localization for mobile robots, Robotics and Autonomous Systems 25 (3–4) (1998) 195–207.
[24] D. Fox, W. Burgard, S. Thrun, Markov localization for mobile robots in dynamic environments, J. Artificial Intelligence Res. 11 (1999) 391–427.
[25] T. Fukuda, S. Ito, N. Oota, F. Arai, Y. Abe, K. Tanake, Y. Tanaka, Navigation system based on ceiling landmark recognition for autonomous mobile robot, in: Proc. Internat. Conference on Industrial Electronics Control and Instrumentation (IECON'93) Vol. 1, 1993, pp. 1466–1471.
[26] W.R. Gilks, S. Richardson, D.J. Spiegelhalter (Eds.), Markov Chain Monte Carlo in Practice, Chapman and Hall/CRC, 1996.
[27] J.-S. Gutmann, W. Burgard, D. Fox, K. Konolige, An experimental comparison of localization methods, in: Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-98), Victoria, BC, 1998.
[28] J.-S. Gutmann, C. Schlegel, Amos: Comparison of scan matching approaches for self-localization in indoor environments, in: Proc. 1st Euromicro Workshop on Advanced Mobile Robots, IEEE Computer Society Press, 1996.
[29] D. Heckerman, A tutorial on learning with bayesian networks, Technical Report MSR-TR-95-06, Microsoft Research, 1995, revised November 1996.
[30] J. Hertzberg, F. Kirchner, Landmark-based autonomous navigation in sewerage pipes, in: Proc. 1st Euromicro Workshop on Advanced Mobile Robots, 1996, pp. 68–73.
[31] R. Hinkel, T. Knieriemen, Environment perception with a laser radar in a fast moving robot, in: Proc. Symposium on Robot Control, Karlsruhe, Germany, 1988, pp. 68.1–68.7.
[32] M. Isard, A. Blake, Contour tracking by stochastic propagation of conditional density, in: Proc. European Conference on Computer Vision, Cambridge, UK, 1996, pp. 343–356.
[33] M. Isard, A. Blake, Condensation: Conditional density propagation for visual tracking, Internat. J. Comput. Vision 29 (1) (1998) 5–28.
[34] P. Jensfelt, S. Kristensen, Active global localisation for a mobile robot using multiple hypothesis tracking, in: Proc. IJCAI Workshop on Reasoning with Uncertainty in Robot Navigation, Stockholm, Sweden, 1999, pp. 13–22.
[35] S.J. Julier, J.K. Uhlmann, A new extension of the kalman filter to nonlinear systems, in: Proc. AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls, 1997.
[36] L.P. Kaelbling, A.R. Cassandra, J.A. Kurien, Acting under uncertainty: Discrete Bayesian models for mobile-robot navigation, in: Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-96), Osaka, Japan, 1996.
[37] R.E. Kalman, A new approach to linear filtering and prediction problems, Trans. ASME, J. Basic Engineering 82 (1960) 35–45.
[38] K. Kanazawa, D. Koller, S.J. Russell, Stochastic simulation algorithms for dynamic probabilistic networks, in: Proc. 11th Annual Conference on Uncertainty in AI, Montreal, Quebec, 1995.

[39] G. Kitagawa, Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, J. Computational and Graphical Statistics 5 (1) (1996) 1–25.
[40] S. Koenig, R. Simmons, Passive distance learning for robot navigation, in: L. Saitta (Ed.), Proc. 13th International Conference on Machine Learning, Bari, Italy, 1996.
[41] D. Kortenkamp, R.P. Bonasso, R. Murphy (Eds.), AI-based Mobile Robots: Case Studies of Successful Robot Systems, MIT Press, Cambridge, MA, 1998.
[42] A.T. Le, Q. Nguyen, Q.P. Ha, D.C. Rye, H.F. Durrant-Whyte, M. Stevens, V. Boget, Towards autonomous excavation, in: A. Zelinsky (Ed.), Field and Service Robotics, Springer, London, 1998, pp. 124–129.
[43] S. Lenser, M. Veloso, Sensor resetting localization for poorly modelled mobile robots, in: Proc. IEEE International Conference on Robotics and Automation (ICRA-2000), San Francisco, CA, 2000.
[44] J.J. Leonard, H.F. Durrant-Whyte, Directed Sonar Sensing for Mobile Robot Navigation, Kluwer Academic, Boston, MA, 1992.
[45] J.J. Leonard, H.F. Durrant-Whyte, I.J. Cox, Dynamic map building for an autonomous mobile robot, Internat. J. Robotics Res. 11 (4) (1992) 89–96.
[46] J. Liu, R. Chen, Sequential Monte Carlo methods for dynamic systems, J. Amer. Statist. Assoc. 93 (1998) 1032–1044.
[47] F. Lu, E. Milios, Globally consistent range scan alignment for environment mapping, Autonomous Robots 4 (1997) 333–349.
[48] J. MacCormick, A. Blake, A probabilistic exclusion principle for tracking multiple objects, in: Proc. International Conference on Computer Vision, Kerkyra, Korfu, 1999.
[49] R.E. Maeder, Ray tracing and graphics extensions, Math. J. 4 (3) (1994).
[50] S. Mahadevan, N. Khaleeli, Robust mobile robot navigation using partially-observable semi-Markov decision processes, Internal Report, 1999.
[51] P. Maybeck, Stochastic Models, Estimation, and Control, Vol. 1, Academic Press, New York, 1979.
[52] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, Wiley Series in Probability and Statistics, Wiley, New York, 1997.
[53] A.W. Moore, Efficient memory-based learning for robot control, Ph.D. Thesis, Trinity Hall, University of Cambridge, Cambridge, 1990.
[54] H.P. Moravec, Sensor fusion in certainty grids for mobile robots, AI Magazine 9 (2) (1988) 61–74.
[55] I. Nourbakhsh, R. Powers, S. Birchfield, DERVISH—An office-navigating robot, AI Magazine 16 (2) (1995) 53–60.
[56] S. Oore, G.E. Hinton, G. Dudek, A mobile robot that learns its place, Neural Computation 9 (1997) 683–699.
[57] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.
[58] M. Pitt, N. Shephard, Filtering via simulation: Auxiliary particle filter, J. Amer. Statist. Assoc. 94 (1999) 590–599.
[59] W.D. Rencken, Concurrent localisation and map building for mobile robots using ultrasonic sensors, in: Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-93), Yokohama, Japan, 1993, pp. 2129–2197.
[60] J. Reuter, Mobile robot self-localization using PDAB, in: Proc. IEEE International Conference on Robotics and Automation (ICRA-2000), San Francisco, CA, 2000.
[61] S.I. Roumeliotis, G.A. Bekey, Bayesian estimation and Kalman filtering: A unified framework for mobile robot localization, in: Proc. IEEE International Conference on Robotics and Automation (ICRA-2000), San Francisco, CA, 2000, pp. 2985–2992.
[62] D.B. Rubin, Using the SIR algorithm to simulate posterior distributions, in: M.H. Bernardo, K.M. DeGroot, D.V. Lindley, A.F.M. Smith (Eds.), Bayesian Statistics 3, Oxford University Press, Oxford, UK, 1988.
[63] S. Scheding, E.M. Nebot, M. Stevens, H.F. Durrant-Whyte, Experiments in autonomous underground guidance, in: R. Harrigan, M. Jamshidi (Eds.), Proc. IEEE International Conference on Robotics and Automation (ICRA-97), Albuquerque, NM, 1997, pp. 1898–1903.
[64] B. Schiele, J. Crowley, A comparison of position estimation techniques using occupancy grids, in: Proc. 1994 IEEE International Conference on Robotics and Automation (ICRA-94), San Diego, CA, 1994, pp. 1628–1634.
[65] R. Simmons, R. Goodwin, K. Haigh, S. Koenig, J. O'Sullivan, A layered architecture for office delivery robots, in: Proc. 1st International Conference on Autonomous Agents, Marina del Rey, CA, 1997.

[66] R. Simmons, S. Koenig, Probabilistic robot navigation in partially observable environments, in: Proc. IJCAI-95, Montreal, Quebec, 1995, pp. 1080–1087.
[67] A.F.M. Smith, A.E. Gelfand, Bayesian statistics without tears: A sampling-resampling perspective, American Statistician 46 (2) (1992) 84–88.
[68] R. Smith, M. Self, P. Cheeseman, Estimating uncertain spatial relationships in robotics, in: I.J. Cox, G.T. Wilfong (Eds.), Autonomous Robot Vehnicles, Springer, Berlin, 1990, pp. 167–193.
[69] M.A. Tanner, Tools for Statistical Inference, 3rd edn, Springer, New York, 1996.
[70] S. Thrun, Bayesian landmark learning for mobile robot localization, Machine Learning 33 (1) (1998).
[71] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Hähnel, C. Rosenberg, N. Roy, J. Schulte, D. Schulz, Probabilistic algorithms and the interactive museum tour-guide robot Minerva, Internat. J. Robotics Res. 19 (11) (2000) 972–999.
[72] S. Thrun, D. Fox, W. Burgard, Monte Carlo localization with mixture proposal distribution, in: Proc. AAAI-2000, Austin, TX, 2000.
[73] M. Vukobratović, Introduction to Robotics, Springer, Berlin, 1989.
[74] G. Weiß, C. Wetzler, E. von Puttkamer, Keeping track of position and orientation of moving indoor systems by correlation of range-finder scans, in: Proc. International Conference on Intelligent Robots and Systems (IROS-94), Munich, Germany, 1994, pp. 595–601.
[75] B. Yamauchi, R. Beer, Spatial learning for navigation in dynamic environments, IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics (Special Issue on Learning Autonomous Robots) (1996), also located at http://www.aic.nrl.navy.mil/~yamauchi/.
[76] S. Zilberstein, S. Russell, Approximate reasoning using anytime algorithms, in: S. Natarajan (Ed.), Imprecise and Approximate Computation, Kluwer Academic, Dordrecht, 1995.