The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

An Application of Kullback-Leibler Divergence to Active SLAM and Exploration with Particle Filters

Luca Carlone, Jingjing Du, Miguel Kaouk Ng, Basilio Bona, and Marina Indri

Abstract— **Autonomous exploration under uncertain robot position requires the robot to plan a suitable motion policy in order to visit unknown areas while minimizing the uncertainty on its pose. The corresponding problem, namely *active SLAM (Simultaneous Localization and Mapping) and exploration* has received a large attention from the robotic community for its relevance in mobile robotics applications. In this work we tackle the problem of active SLAM and exploration with Rao-Blackwellized Particle Filters. We propose an application of Kullback-Leibler divergence for the purpose of evaluating the particle-based SLAM posterior approximation. This metric is then applied in the definition of the *expected gain from a policy*, which allows the robot to autonomously decide between exploration and place revisiting actions (i.e., loop closing). The technique is shown to enhance robot awareness in detecting loop closing occasions, which are often missed when using other state-of-the-art approaches. Results of extensive tests are reported to support our claims.**

I. INTRODUCTION

Navigation capabilities represent a crucial prerequisite for a mobile robot to perform tasks in unknown and potentially unstructured environments. The increasing request of autonomy further stresses the importance of overcoming purely reactive behaviors, in order to enhance the decision making capability of a robot exploring the environment. When prior knowledge of robot position is available, solutions which work reasonably well for exploration do exist, and the robot decision reduces to the choice of an exploration target which maximizes the opportunity of visiting unknown areas. However, when the more general problem of exploration under uncertain robot position is tackled, several challenges arise. The corresponding problem, also referred to as *active SLAM and exploration*, requires the robot to actively control its movements in order to maximize the explored areas and at the same time minimize the uncertainty in Simultaneous Localization and Mapping. The former objective is clearly accomplished by visiting unknown places, whereas the latter requires the robot to perform loop closing actions, that is to come back to already traveled areas. In order to properly plan its policy, the robot needs reliable metrics for trading-off between the two action typologies, evaluating the expected gain of a motion policy.

When dealing with landmark-based representations, this metric can be naturally found in the entropy of the multi variate belief describing both map and robot pose. Although the computation of the expected gain of an action can be computational demanding, effective and well founded approaches exist, see for instance [1], [2], [3], and the references therein. The problem of active SLAM and exploration with metric representations, instead, is not fully understood and the recent literature on the topic remarks several drawbacks of naive entropy-based metrics [4]. In this last scenario Rao-Blackwellized Particle Filters (RBPF) have been demonstrated to be an effective solution for the estimation of high resolution world models, in the form of occupancy grid maps. The application of particle filters to SLAM is based on Rao-Blackwell factorization: the particle filter is applied to the problem of estimating potential trajectories and a map hypothesis is associated to each sample. Although it is possible to guarantee (almost sure) convergence of the particle-based posterior toward the true one for a particle set size going to infinite [5], a common processing unit is not able to deal with a particle set larger than few hundreds of samples. As a consequence, the less particles are available the worse is the approximation of the true posterior, and this issue becomes critical when the amount of uncertainty in the filter increases. From this consideration it stems the importance of loop closing in RBPF-SLAM: place revisiting actions lead to a reduction of uncertainty and hence improve filter capability in modeling the true posterior.

In this work we investigate the problem of active SLAM and exploration with Rao-Blackwellized Particle Filters. We first derive a measure of RBPF uncertainty by using the Kullback-Leibler divergence [6] for trajectory belief evaluation. The divergence allows to obtain an upper bound on the error when approximating the true posterior with a particle-based belief representation. Then this metric is successfully employed in the definition of the *expected information from a policy*. It is worth noticing that the term "expected" is used in the last sentence in the probabilistic sense and not only to resemble the predictive nature of the metric. The probabilistic interpretation of the information gain we propose is then validated by comparing it with three metrics of information gain, namely a *naive gain*, an *entropy-based gain* [7], and the *expected map information* [4]. We further validate our approach by considering the case of a robot deployed in an unknown environment, which has to perform autonomous exploration exploiting the information gain computed with our technique.

The paper is organized as follows. Related works on active SLAM and exploration are reviewed in Section II. Kullback-Leibler divergence is applied for SLAM posterior evaluation in Section III-A, then its application to information gain-based exploration is presented in Section III-B. Numerical experiments are reported in Section IV, while conclusions are drawn in Section V.

This work was supported in part by Regione Piemonte under MACP4log Grant (RU/02/26).

L. Carlone and M. Kaouk Ng are with CSPP, Laboratorio di Meccatronica, Politecnico di Torino, Torino, Italy, {luca.carlone, miguel.kaoukng}@polito.it

J. Du, B. Bona and M. Indri are with Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, Italy, {jingjing.du, basilio.bona, marina.indri}@polito.it

978-1-4244-6676-4/10/$25.00 ©2010 IEEE

II. Related Work

In this section we discuss the state-of-the-art of robotic exploration using occupancy grid map representation [8]. The literature on robotic exploration can be classified in two wide research lines: exploration with perfect position knowledge and exploration under uncertain localization. In the former scenario, accurate prior knowledge of robot position is available and the robot decision making reduces to the choice of the exploration target that maximizes the opportunity of visiting unknown areas. Early contribution to this framework can be found in [9], in which *frontiers*, that is cells on the boundary between known areas and unexplored space, are selected and used as potential targets for exploration. In such approaches the robot simply chooses the nearest target, regardless complex decision making strategies. Similar approaches can be found in recent contributions, see for instance [10]. Also in the case of perfect knowledge of robot pose, the onboard sensors are affected by uncertainty, hence suitable metrics are required for evaluating the so called *information gain* that is expected when choosing a target for exploration. In order to model the uncertainty of the occupancy grid map representation of the environment, Bourgault et al. [1] and Moorehead et al. [11] proposed the use of the entropy of the cells in the grid. In such a case the gain, computed from the entropy, quantifies the expected amount of information that can be acquired by taking a sensor reading in a certain area of the map. In [12] an entropy-based gain is employed for target selection, but it is computed on an augmented representation of the environment, referred to as *coverage map*.

The approaches mentioned so far do not take into account the pose uncertainty when selecting the next vantage point. When dealing with the general problem of exploration under uncertain robot position, several challenges arise. Stachniss et al. addressed the problem of active SLAM and exploration using Rao-Blackwellized Particle Filters, improving exploration performances via active loop closing, through a heuristic method for re-traverse loops [13]. A decision-theoretic approach is proposed in [7], in which active loop closing is achieved by monitoring the uncertainty in RBPF. Many drawbacks of such method are analyzed in [4], where the authors also proposed a novel measurement of RBPF-SLAM uncertainty. The importance of uncertainty metrics stems from the fact that they allow to discern possible loop closing actions when performing exploration. Details on related techniques are further discussed in Section IV-A.

III. Active SLAM and Exploration

A. Rao-Blackwellization and Map Consistency

The high dimensionality of state space in grid-based SLAM makes challenging the application of sample-based representations of the posterior of robot pose and occupancy grid map. An elegant solution to reduce dimensionality of the sampling space can be obtained through Rao-Blackwellization [14]. Since the map probability can be computed analytically given the robot path, it is possible to factorize the joint probability as follows:

$$p(x_{1:t}, m \mid z_{1:t}, u_{0:t-1}) = p(m \mid x_{1:t}, z_{1:t}) \cdot p(x_{1:t} \mid z_{1:t}, u_{0:t-1}) \quad (1)$$

In (1) the state includes the robot trajectory $x_{1:t} = \{x_1, x_2, \ldots, x_t\}$ and the map m, both estimated from the measurements $z_{1:t} = \{z_1, z_2, \ldots, z_t\}$ and the commands $u_{0:t-1} = \{u_0, u_1, \ldots, u_{t-1}\}$. Eteroceptive measurements and odometry are often included in the data vector $d_{1:t} = \{z_{1:t}, u_{0:t-1}\}$. Equation (1) provides the basis for Rao-Blackwellized Particle Filters SLAM: the particle filter is applied to the problem of estimating potential trajectories and a map hypothesis is associated to each sample. Due to memory constrains, the filter risks to become inconsistent when the uncertainty about pose estimation grows so much that it cannot be properly modeled by the particle-based belief. Therefore the loop closing becomes crucial, since place revisiting actions lead to a reduction of uncertainty and hence improve filter capability in modeling the true posterior. Similar observations were formalized by Fox [15] using Kullback-Leibler divergence, with application to Monte Carlo Localization. Kullback-Leibler divergence is a common measure of fit between two probability distributions, in our case the true posterior and the point mass approximation. Starting from the assumption that the true posterior is a discrete piecewise constant distribution, with support on finite patches (later referred to as *bins*), Fox derived the number of particles needed to achieve a desired approximation error with a given probability.

Conversely we now consider a fixed particle set size n and we evaluate the error between the true posterior $p(x_t \mid d_{1:t})$, describing robot pose at time t, and its point mass approximation $\hat{p}(x_t \mid d_{1:t})$. Starting from [15], with simple computation, we can derive the following upper bound on the divergence between the true and approximated pose belief:

$$\xi(\hat{p}(x_t \mid d_{1:t}), p(x_t \mid d_{1:t})) < \bar{\xi},$$
$$\bar{\xi} = \frac{k_t - 1}{2n}\left[1 - \frac{2}{9(k_t-1)} + \sqrt{\frac{2}{9(k_t-1)}}\, z_{1-\delta_t}\right]^3, \quad (2)$$

which holds with probability $1 - \delta_t$. In (2) k_t is the number of bins, being support of the true discrete distribution, in which at least one particle falls, and $z_{1-\delta_t}$ is the upper $1 - \delta_t$ quantile of the standard normal distribution.

We now provide an application of the methodology proposed by Fox for on-line evaluation of SLAM posterior approximation. It is worth underlining that, while in localization the posterior approximation influences robot pose estimation performance, in Simultaneous Localization and Mapping an inadequate filter approximation can compromise the whole mapping process, leading the robot to fail in accomplishing consistent mapping. According to Blackwellization the probability $p(m \mid x_{1:t}, z_{1:t})$ can be computed analytically given the trajectory, hence treats to filter consistency have to be found in the estimation of robot poses. For sake of simplicity, in the following we write the divergence as $\xi(p(x_t \mid d_{1:t}))$, omitting the dependence on the two compared distributions. According to (2), for each pose $x_i, i = 1, 2, ..., t$ we can define an upper bound $\bar{\xi}_i$ on the approximation error with a given probability $1 - \delta$. Let the error of approximation of trajectory posterior be defined as:

$$\xi(p(x_{1:t}, d_{1:t})) = \xi(p(x_t^*, d_{1:t})), \quad (3)$$
$$x_t^* = \underset{i \in [1:t], 1-\delta_i = 1-\delta}{\arg\max} (\bar{\xi}_i)$$

that is we assign to $\xi(p(x_{1:t}, d_{1:t}))$ the error of the pose which is expected to provide the worst approximation. Equation (3) formalizes the observation that the trajectory estimation, thus the map quality, can be compromised if also a single pose is affected by large errors. The relevance of pose description in SLAM posterior estimation is remarked in [16], where similar conclusions are applied for the purpose of benchmarking different SLAM approaches. According to equation (3), and since $\bar{\xi}_i$ is monotonically increasing in the number of non-empty bins, we can derive the probability of having a trajectory error less than $\bar{\xi}$ as:

$$p(\xi(p(x_{1:t}, d_{1:t})) < \bar{\xi}) = p(\xi(p(x_t^*, d_{1:t})) < \bar{\xi}) = \qquad (4)$$

$$= \mathcal{F}\left\{\left[\sqrt{\frac{9(k^*-1)}{2}} - \left(\sqrt[3]{\frac{2n\bar{\xi}}{k^*-1}} + \frac{2}{9(k^*-1)} - 1\right)\right]\right\},$$

where $k^* = \max\{k_i, i \in [1:t]\}$ and $\mathcal{F}(\cdot)$ is the cumulative distribution function of a standard Gaussian distribution. In the rest of this section we discuss how to properly select the parameters of the proposed model so that the probability $p(\xi(\hat{p}(x_{1:t}, d_{1:t})) < \bar{\xi})$ actually reflects the probability of estimating a consistent map. With a given particle set size, equation (4) depends on the number of non-empty bins, k_t, and the error bound $\bar{\xi}$. The former is connected with the bin size we select for state space discretization. In our previous work [17] we extensively tested different bin sizes and we provided a simple procedure to properly choose this parameter. Roughly speaking the variable k_t ranges from one to n (number of particles) and the bin size has to be designed to obtain an outcome of the model as informative as possible: extremely small grid sizes, lead to fast saturation of k_t (particles cover all possible bins in few iterations of the filter), whereas large grid sizes do not give a fine granularity in uncertainty evaluation (particles always cover few bins, hence only small values of k_t occur). Our investigation leads us to use bin sizes comparable with the dimension of the cells in the occupancy grid map for the robot position, whereas the orientation is discretized with resolution of 5^o.

The error bound $\bar{\xi}$, instead, relates the number of non-empty bins with the probability of satisfying the inequality between the trajectory approximation error and its upper bound. For Rao-Blackwell factorization inconsistencies in trajectory estimation are unavoidably reflected in incorrect map estimation. As a consequence we can just select $\bar{\xi}$, so that the probability (4) corresponds the probability of estimating a consistent map. Although in grid-based SLAM visual inspection is a common approach for evaluating map quality, we preferred to give quantitative evidence of the estimated map quality by comparing it with the corresponding ground truth (available in simulation). For this purpose we used the metric proposed by Carpin in [18] and reported in the following definition.

Definition 1: Let M_1 and M_2 be two grid maps. The agreement between M_1 and M_2 (indicated as $agr(M_1, M_2)$) is the number of cells in M_1 and M_2 that are both free or both occupied. The disagreement between M_1 and M_2 (indicated as $dis(M_1, M_2)$) is the number of cells such that M_1 is free and M_2 is occupied or vice-versa. The *acceptance index* between them is defined as:

$$\omega(M_1, M_2) = \begin{cases} 0 & if\ agr(M_1, M_2) = 0 \\ \frac{agr(M_1, M_2)}{agr(M_1, M_2) + dis(M_1, M_2)} & if\ agr(M_1, M_2) \neq 0 \end{cases}$$

The acceptance index, ranging between 0 and 1, gives information on map similarity, once a suitable roto-translation is applied. We performed extensive tests on RBPF-SLAM

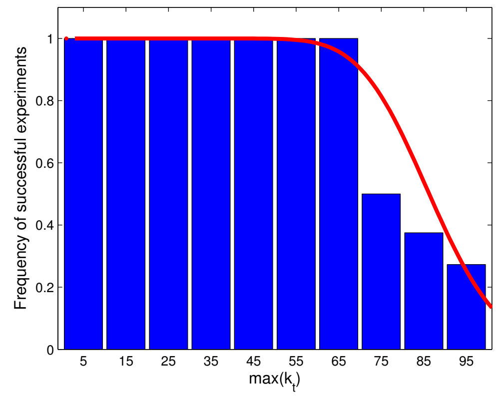

Fig. 1. Histogram represents the frequency of successful experiments for each maximum observed value of k_t. Tests were performed with a particle set size $n = 100$.

and for each run we recorded the maximum k_t and the corresponding acceptance index. If the latter is higher than a threshold the map is classified as consistent (successful experiment), otherwise the map building process failed (in our tests the threshold was fixed to 0.85 although the classification showed low sensitivity on this parameter). The histogram in Figure 1 describes the outcome of our tests: for each maximum observed k_t, we associate the frequency of successful experiments, that is the ratio between the number of experiments in which a consistent map is estimated and the total number of experiments in which the same maximum k_t is observed. Therefore we can simply compute $\bar{\xi}$ so that the probability obtained from (4) corresponds to the experimental data contained in the histogram. In our implementation we used a least-squares fitting and we obtained a value $\bar{\xi} = 0.42$ (the corresponding curve is shown in red in Figure 1). Notice that, if a more conservative metric is needed, the $\bar{\xi}$ can be selected so that the red curve is always below the histogram of the experimental data. As discussed in Section IV, this parameter provides a degree of freedom in controlling the robot behavior, hence being useful for tailoring the exploration strategy on the particular robotic application. If a lower value of $\bar{\xi}$ is selected the robot will choose vantage points that minimize the risk of loosing information (or, equivalently, to minimize the uncertainty in the filter), otherwise, with a higher value of this parameter, it will prefer to maximize the possibility of visiting new places. In such a way $\bar{\xi}$ is crucial in defining a suitable tradeoff between active SLAM and exploration.

B. The Expected Information from a Policy

The previous section highlighted a relevant aspect of map building using Rao-Blackwellized Particle Filter. For the nature of the approximation of SLAM posterior the filter is able to assure consistent mapping only under correct trajectory approximation. As a consequence the concept of *expected information gain* has to be revisited: classic information gain metrics quantify the amount of information introduced in the filter, but do not take into account the possibility of information loss due to inconsistency in the filter. The robot, while exploring new areas, gains information, but at the same time it risks an information loss since the filter approximation worsens. In Section III-A we computed a probability that quantifies the risk of incorrect trajectory approximation, hence we can formulate the expected information gain as follows.

Definition 2: Let $I(m_t) = I\left(p(m|x_{1:t}, d_{1:t})\right)$ be the current map information, available to the robot at time t. Assume that a reasonable prediction of the amount of information, which can be acquired when applying the motion policy π_i for reaching the exploration target i, can be computed. If we call $I(m_{t+T(\pi_i)}) = I\left(p(m|x_{1:t+T(\pi_i)}, d_{1:t+T(\pi_i)})\right)$ the predicted information after the target is reached, the *expected information from the policy* π_i is defined as:

$$\begin{aligned} E[I(\pi_i)] = p(t,\pi_i)[I(m_{t+T(\pi_i)}) - I(m_t)] + \\ +(1-p(t,\pi_i))[-I(m_t)] \end{aligned} \tag{5}$$

where $p(t,\pi_i) = p(\xi(p(x_{1:t+T(\pi_i)}, d_{1:t+T(\pi_i)}) < \bar{\xi})$.

The previous quantity is simply the expected value over the possible results of policy application: with probability $p(t,\pi_i)$ the estimated map is consistent, hence the robot has the possibility to gain the information $[I(m_{t+T(\pi)}) - I(m_t)]$, whereas with probability $1 - p(t,\pi_i)$ the filter becomes inconsistent and the robot can no longer model SLAM posterior in a proper way, loosing the information acquired before time t. When the robot is confident on its SLAM approximation ($p(t,\pi_i) \approx 1$), equation (5) is dominated by the first summand, hence the robot will prefer to maximize the gain $I(m_{t+T(\pi_i)})$ selecting exploration actions (move towards unknown areas). When the probability $p(t,\pi_i)$ drops, that is the robot has traveled in unknown areas and the uncertainty in its approximation is getting worse, its policy will tradeoff between loop closing actions (which contribute to $E[I(\pi_i)]$ with an increase of $p(t,\pi_i)$) and exploration actions (which contribute to $E[I(\pi_i)]$ with an increase of $I(m_{t+T(\pi_i)})$). Of course the expected gain in some targets can result in a negative number, truly mirroring the risk of information loss. It is worth noticing that we provide no specification on how compute the information I, in order to maintain the generality of the approach. Common entropy-based information metrics can be used whereas in the following section we consider an even simpler measure, which still preserves the desirable characteristics of our approach.

IV. Experiments

In this section, we present the results of the implementation of our method and compare it with the performances of other related techniques, namely an *entropy-based* gain [7], the *expected map information* gain [19], [4] and a simpler metric (*naive* gain). The tests were designed in MobileSim and the robotic platform used was an *ActivMedia Pioneer P3-DX* equipped with a laser range sensor and odometry pose estimation. In our RBPF-SLAM implementation, we incorporate two important aspects such as the *adaptive resampling technique* [20] to reduce the problem of particle depletion, and the *laser-stabilized odometry* [21] for improving the prediction phase of the filter.

A. Compared Techniques

The first metric we consider is an *entropy-based* gain, $G[H]$, computed from the *joint entropy* of robot poses and occupancy grid map. It is defined as the expected joint entropy reduction when the robot execute an action:

$$\begin{aligned} G_i[H] \;=\; H\left(p(x_{1:t},\, m \mid d_{1:t})\right) \;+ \\ -\, H\left(p(x_{1:t+T(\pi_i)},\, m \mid d_{1:t+T(\pi_i)})\right) \end{aligned} \tag{6}$$

where the first term is the joint entropy of the particle filter, computed at time t, and the second term is the predicted joint entropy after the target i is reached (following the policy π_i). According to Stachniss et al. [7] the computation of the *joint entropy* of a RBPF can be approximated as:

$$\begin{aligned} H\left(p(x_{1:t}, m \mid d_{1:t})\right) \;\approx\; H\left(p(x_{1:t} \mid d_{1:t})\right) + \\ + \textstyle\sum_{j=1}^{n}\; w_t^{[j]}\, H\left(p(m^{[j]} \mid x_{1:t}^{[j]},\, d_{1:t})\right) \end{aligned} \tag{7}$$

where $w_t^{[j]}$ is the weight of the *j-th* particle at time step t, $x_{1:t}^{[j]}$ and $m^{[j]}$ are the trajectory and the map associated with the *j-th* particle. The first term of (7) represents the entropy of trajectory posterior, whereas the second one corresponds to a weighted average of individual map entropies. Entropy of trajectory posterior can be approximated as the entropy of the Gaussian distribution (in $\mathbb{R}^{3\times t}$) fitting particle poses or using other approximations. Roy et al. [22] and subsequently Stachniss et al. [7] used the following simplification for reducing the computational effort:

$$H\left(p(x_{1:t} \mid d_{1:t})\right) \approx \frac{1}{t}\sum_{\tau=1}^{t}\; H\left(p(x_\tau \mid d_{1:\tau})\right) \tag{8}$$

where $p(x_\tau \mid d_{1:\tau})$ is a Gaussian approximation (in $\mathbb{R}^3$) of pose posterior at time τ.

The second information gain is based on the computation of the *Expected Map Information* (EMI). The Expected Map (EM) is a grid map in which the occupancy of each cell $p(EM_{xy} \mid d_{1:t})$ considers the contribution of all the particles that compound the set:

$$p(EM_{xy} \mid d_{1:t}) \approx \sum_{j=1}^{n} w_t^{[j]} p(m_{xy}^{[j]} \mid x_{1:t}^{[j]},\, d_{1:t}) \tag{9}$$

Therefore the information of the expected map $I\left(p(EM \mid d_{1:t})\right)$ can be computed as:

$$\begin{aligned} I\left(p(EM \mid d_{1:t})\right) &= \textstyle\sum_{\forall x,y} I(p(EM_{xy} \mid d_{1:t})) \\ I\left(p(EM_{xy} \mid d_{1:t})\right) &= 1 \;-\; H(p(EM_{xy} \mid d_{1:t})), \end{aligned} \tag{10}$$

where $H\left(p(EM_{xy} \mid d_{1:t})\right)$ is the entropy of a grid cell in the expected map. Therefore, we calculate the *Expected Map Information* gain $G[EMI]$ as the difference between the predicted information in the EM when the robot applies the policy π_i towards the exploration target i and the EMI computed at time t:

$$G_i[EMI] = I\left(p(EM \mid d_{1:t+T(\pi_i)})\right) - I\left(p(EM \mid d_{1:t})\right) \tag{11}$$

The third metric that is taken into account is based on the number of cells in the occupancy grid map of the best sample (particle with highest weight) that can be visited when performing a policy. It is called *naive* since it neglects both the uncertain description of robot poses and the probabilistic interpretation of sensor measurements. We consider a cell as observed when its occupancy likelihood is different from 0.5. Starting from this consideration, we define the *Naive* expected gain $G[N]$ as the difference between the predicted number of cells observed when the target i is reached and the number of cells observed in the current pose of the robot at time t:

$$G_i[N] \ = \ N_{t+T(\pi_i)} - \ N_t \tag{12}$$

Finally we computed our metric using the number of observed cells N_t as information measure $I(m_t)$. In this case, with simple computation from (5), our *expected information from the policy* π_i reduces to:

$$G_i[EI] = p(t, \pi_i) N_{t+T(\pi_i)} - N_t \tag{13}$$

We remark that, when computing the information gains for each target, according to (6), (11), (12) and (13), the terms referred to the current information (at time t) constitute an offset which is equal for all targets, hence we neglect them in the following.

B. Results and Discussion

We first consider some relevant case studies in which the robot is called to evaluate the information gain at some possible target points. In a further experiment the $G[EI]$ is actually used in a simple experiment of autonomous exploration.

As reported in [4] and [7] the predicted laser measurements (from the current position to the target) are obtained with a ray-casting operation on the map of the best particle. The same map is used to plan the path to reach the target. In our implementation path planning is performed using the A^* algorithm. Using the estimated sensor observation the robot can run a simulation of the particle filter, for the purpose of computing the expected gain in each target. In our experiments unknown areas are supposed to be obstacle free although different assumption can be considered when prior knowledge on the environment is available.

*1) **Case studies**:* We first consider a scenario in which the robot starts to navigate in the environment and encounters a crossroad, as shown in Figure 2. In such a case the robot has four potential targets: target 4 corresponds to a place revisiting action whereas the others lead robot to explore new areas. The four subplots in the figure show the simulation of the particle filter behavior for each target. Figure 2 also reports the value of the information gains for each of the techniques introduced so far. In this example all the techniques lead to the same decision: at the instant of decision making the robot does not see both walls of the side corridors, hence it expects an higher gain (higher information or, dually, smaller entropy) in moving towards targets 1 and 3. This result resembles human behavior in that it leads the explorer to gain more information by taking a look at both corridors, instead of blindly proceeding straight on.

The second case study reports a loop closing scenario, similar to the one tested in [4] and [13]. The test environment

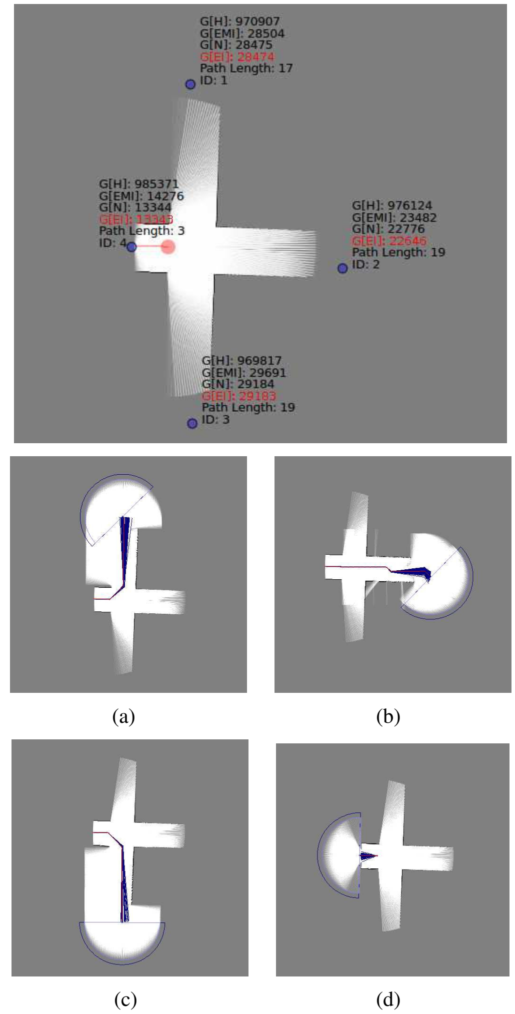

Fig. 2. Case study 1: Crossroad. Robot is shown as a red dot (traveled trajectory is shown in red too). For each target we report an identification number (ID), the computed information gains and the estimated path length to reach the target. (a), (b), (c), and (d) show the predicted behavior of the particle filter for each target.

is shown in Figure 3 whereas the estimated gains for each target are reported in Figure 4 (for targets 4 and 9 no path was found by the A^* algorithm since they are too close to an obstacle). As a result we notice that the gain computed from the number of visited cells ($G[N]$) would lead the robot towards target 5, preferring an exploration action. This behavior was expected since $G[N]$ does not take into account robot uncertainty. Also for $G[H]$ the target of choice is 5, since it minimizes the expected entropy of the filter. A deeper analysis reveals that this result is common for $G[H]$ for two main reasons: (i) path entropy contribution is negligible with respect to map entropy, which has a commanding influence on $G[H]$; (ii) sample map entropies, with respect to (7), are almost equal, hence the following simplifications are suggested from experimental evidence:

$$\begin{aligned} &\textstyle\sum_{j=1}^{n} \ w_t^{[j]} \ H\left(p(m^{[j]} \mid x_{1:t}^{[j]}, \ d_{1:t})\right) \approx \\ &H\left(p(m^{[i]} \mid x_{1:t}^{[i]}, \ d_{1:t})\right) \textstyle\sum_{j=1}^{n} \ w_t^{[j]} = \\ &H\left(p(m^{[i]} \mid x_{1:t}^{[i]}, \ d_{1:t})\right) = f(N_t) \end{aligned}$$

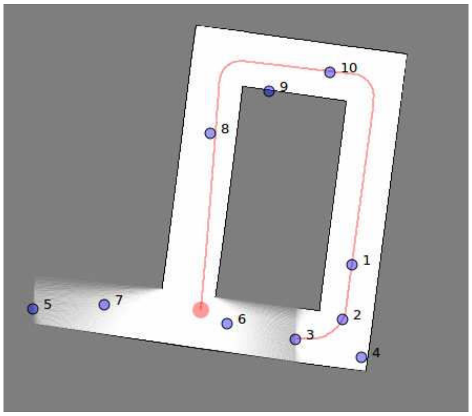

Fig. 3. Case study 2: Loop closing. Robot is shown as a red dot (traveled trajectory is shown in red too).

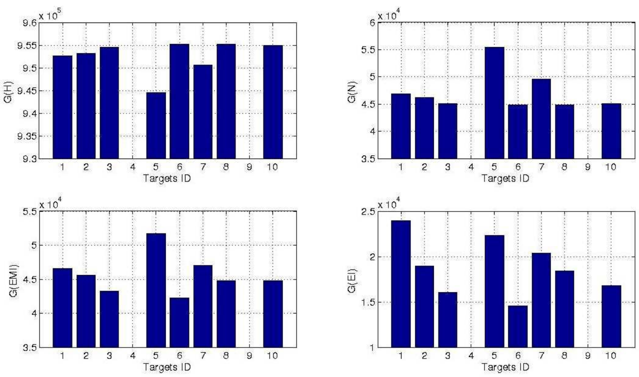

Fig. 4. Bar plot of the information gains for each of the targets in Figure 3. No gain is reported for targets 4 and 9 for which no path was found. Hence the target of choice using $G[H]$, will be the one with the minimum entropy value (lower bar), whereas the other techniques will suggest targets with higher value of the information (higher bar).

where i is the best sample and $f(\cdot)$ is a monotonically decreasing function. Further drawbacks are reported in our previous work [17]. The $G[EMI]$ shows the same undesirable behavior (5 is the target with the highest expected information) and also in this case the metric suggests the robot not to close the loop. In several experiments the gain computed from the expected map information showed no remarkable difference with respect to the one based on joint entropy. Experimental evidence suggests that, although $G[EMI]$ models inconsistency among map hypotheses, it does not provide a reliable tradeoff for robot decision making. Loop closing actions and exploration actions are rewarded by the metric, but in many cases the uncertainty in the particle filter has a minor influence on the expected gain.

Our technique, however, suggests the robot to close the loop, moving towards target 1. Apart from a mere comparison on the numeric outcome of the experiment, we can point out some relevant considerations. On one hand the results enlightens some drawbacks of common frontier-based target selection: reaching the frontier 3 is not sufficient to reduce the amount of uncertainty in the filter. In such a case target 1 is preferred by the robot since the uncertainty decreases only after traveling for a certain time in known areas. The latter is a well known phenomenon, see [7] for instance, and our metric correctly describes it. Moreover, the *expected information from a policy* provides an intuitive way to control robot behavior: if the robot has to be more conservative in exploration we can simply set a lower value of $\bar{\xi}$ in (4). This corresponds to impose a stricter constraint on SLAM uncertainty, hence forcing the robot to prefer loop closing actions instead of visiting unknown areas. If one wants to take the risk of map inconsistency, the upper bound $\bar{\xi}$ is relaxed and the tradeoff will lead towards exploration actions. When $\bar{\xi}$ is further increased, the probability $p(t, \pi_i)$ remains to 1, hence reducing our approach to the naive gain (12).

*2) **Autonomous exploration**:* Here we report a simple experiment in which our *expected gain from the policy* is used for active SLAM and exploration. It is worth noticing that the motion strategy of the robot has to consider both the expected gain from a target and the cost required for reaching it (usually the path length, l_i, to be traveled for reaching the i-th target). Several approaches considered a linear combination of expected gain and cost for the purpose of computing the utility of moving towards a target. Although this definition of utility is well founded and widespread in robotic literature (see [1] and [13]), it requires a careful setting of the coefficient of the linear combination, that has to be done experimentally and can be scenario-dependant. In our implementation we overcome these difficulties by using a *specific (normalized) information gain from the policy*:

$$\rho_i[EI] = \frac{G_i[EI]}{l_i} \tag{14}$$

In the previous expression the expected gain is simply normalized by the distance to be traveled (estimated from A^*-based path planner). Therefore the specific information assumes the meaning of a gain for each meter traveled, and intrinsically takes into account the disadvantage of reaching targets which are far away.

The estimated map in the autonomous exploration experiment is shown in Figure 5. The trajectory followed by the

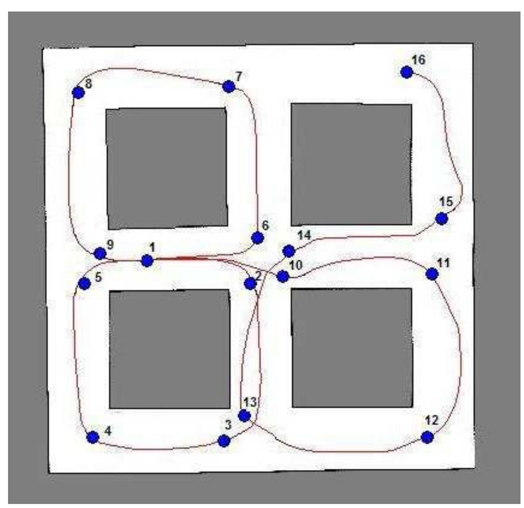

Fig. 5. Autonomous exploration experiment. Motion strategy is computed using the specific information $\rho_i[EI]$, and targets chosen by the robot as best candidate for next motion are displayed. (target 1 is the starting point whereas target 16 is the last position reached).

robot is shown in red, whereas the dots correspond to the targets chosen by the robot as best candidate for the next motion. When a target is reached the robot computes the $\rho_i[EI]$ for all the targets in the map and move towards the target with the highest specific information.

V. Conclusion

In this work we investigated the problem of active SLAM and exploration with Rao-Blackwellized Particle Filters. We proposed an application of Kullback-Leibler divergence for the purpose of evaluating the particle-based SLAM posterior approximation. This metric is then employed in the definition of an *expected gain from a policy*, which allows the robot to autonomously decide the best motion strategy to explore the environment and reduce the uncertainty in SLAM posterior estimation. The probabilistic interpretation of the information gain we propose, is then validated by comparing it with three metrics of information gain, namely a *naive gain*, an *entropy-based gain*, and the *expected map information*. The *expected gain from a policy* is shown to enhance robot awareness in detecting loop closing occasions, which are often missed when using the compared approaches. Current research effort is devoted to test the proposed technique in challenging indoor scenarios and to extend the approach to the multi robot case [23]. Finally, we are studying the possibility of generalizing the theoretical formulation proposed in this paper, in order to cope with the observation that it can be too pessimistic to assume that filter inconsistency causes the loss of all the information acquired by the robot.

References

[1] F. Bourgault, A.A. Makarenko, S.B. Williams, B. Grocholsky, and H.F. Durrant-Whyte. Information based adaptive robotic exploration. In *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002.

[2] R. Martinez-Cantin, O. De Freitas, A. Doucet, and J.A. Castellanos. Active policy learning for robot planning and exploration under uncertainty. In *Proceedings of Robotics: Science and Systems*, 2007.

[3] R. Sim and N. Roy. Global A-optimal robot exploration in SLAM. In *Proceedings of the IEEE lnternational Conference on Robotics and Automation*, pages 673–678, 2005.

[4] J.L. Blanco, J.A. Fernandez-Madrigal, and J. Gonzalez. A novel measure of uncertainty for mobile robot SLAM with Rao-Blackwellized particle filters. *The International Journal of Robotics Research*, 27(1):73–89, 2008.

[5] D. Crisan and A. Doucet. A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on signal processing*, 50(3):736–746, 2002.

[6] S. Kullback and A. Leibler. On information and sufficiency. *Annals of Mathematical statitiscs*, 22:79–86, 1951.

[7] C. Stachniss, G. Grisetti, and W. Burgard. Information gain-based exploration using Rao-Blackwellized particle filters. In *Proceedings of Robotics: Science and Systems*, 2005.

[8] H.P. Moravec. Sensor fusion in certainty grids for mobile robots. *AI Magazine*, 9(2):61–74, 1988.

[9] B. Yamauchi. A frontier-based approach for autonomous exploration. In *Proceedings of CIRA 97*, 1997.

[10] K. Konolige B. Limketkai D. Schulz B. Stewart D. Fox, J. Ko. Distributed multirobot exploration and mapping. *Proceedings of the IEEE*, 94(7):1325–1339, 2006.

[11] S.J. Moorehead, R. Simmons, and W.L. Whitaker. Autonomous exploration using multiple sources of information. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2001.

[12] C. Stachniss and W. Burgard. Exploring unknown environments with mobile robots using coverage maps. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2005.

[13] C. Stachniss, D. Hahnel, and W. Burgard. Exploration with active loop-closing for FastSLAM. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1505–1510, 2004.

[14] A. Doucet, J. de Freitas, K. Murphy, and S. Russel. Rao-Blackwellized particle filtering for dynamic bayesian networks. In *Proceeding of the Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 176–183, 2000.

[15] D. Fox. Adapting the sample size in particle filters through KLD-sampling. *International Journal of Robotics Research*, 22(12):985–1003, 2003.

[16] W. Burgard, C. Stachniss, G. Grisetti, B. Steder, R. Kümmerle, C. Dornhege, M. Ruhnke, A. Kleiner, and Juan D. Tardós. A comparison of SLAM algorithms based on a graph of relations. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009.

[17] L. Carlone, M. Kaouk Ng, J. Du, B. Bona, and M. Indri. Reverse KLD-sampling for measuring uncertainty in rao-blackwellized particle filters SLAM. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009.

[18] S. Carpin. Fast and accurate map merging for multi-robot systems. *Autonomous Robots*, 25:305–316, 2008.

[19] J.L. Blanco, J.A. Fernandez-Madrigal, and J. Gonzalez. An entropy-based measurement of certainty in Rao-Blackwellized particle filter mapping. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robotics and Systems*, pages 3550–3555, 2006.

[20] C. Stachniss, G. Grisetti, D. Hahnel, and W. Burgard. Improved Rao-Blackwellized mapping by adaptive sampling and active loop-closure. In *Proceedings of the Workshop on Self-Organization of AdaptiVE behavior (SOAVE)*, pages 1–15, 2004.

[21] A. Howard. Multi-robot simultaneous localization and mapping using particle filters. *Robotics: Science and Systems*, pages 201–208, 2006.

[22] N. Roy, W. Burgard, D. Fox, and S. Thrun. Coastal navigation – robot motion with uncertainty. 1998.

[23] L. Carlone, M. Kaouk Ng, J. Du, B. Bona, and M. Indri. Rao-Blackwellized particle filters multi robot SLAM with unknown initial correspondences and limited communication. In *Proceedings of the IEEE lnternational Conference on Robotics and Automation*, 2010.